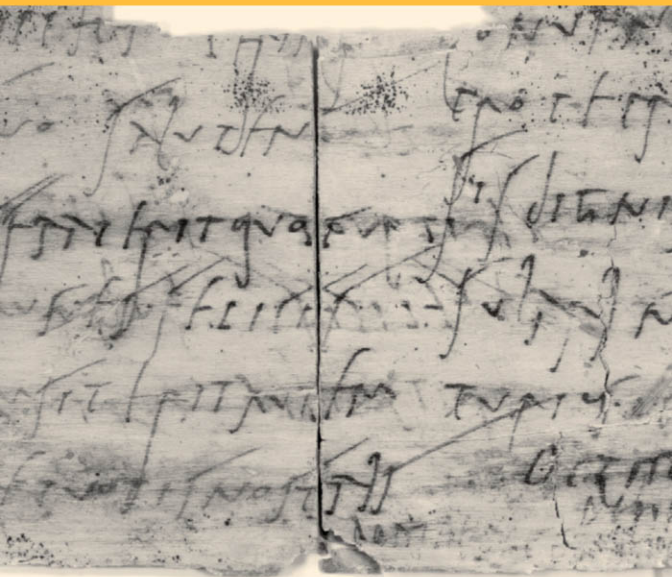


Image to Interpretation

*An Intelligent System to Aid Historians in
Reading the Vindolanda Texts*

Melissa M. Terras

OXFORD STUDIES IN ANCIENT DOCUMENTS



VISIT...

LANZAROTE
Caliente.COM

OXFORD STUDIES IN ANCIENT DOCUMENTS

General Editors

Alan Bowman Alison Cooley

OXFORD STUDIES IN ANCIENT DOCUMENTS

This innovative new series offers unique perspectives on the political, cultural, social, and economic history of the ancient world. Exploiting the latest technological advances in imaging, decipherment, and interpretation, the volumes cover a wide range of documentary sources, including inscriptions, papyri, and wooden tablets.

Image to Interpretation

*An Intelligent System to Aid Historians in
Reading the Vindolanda Texts*

MELISSA M. TERRAS

*With a foreword by Professor Alan K. Bowman and Professor
Michael Brady*

*Chapters 4, 5, and Appendix A co-authored with Dr Paul
Robertson*

OXFORD
UNIVERSITY PRESS

OXFORD

UNIVERSITY PRESS

Great Clarendon Street, Oxford ox2 6DP

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide in

Oxford New York

Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in

Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Oxford is a registered trade mark of Oxford University Press
in the UK and in certain other countries

Published in the United States
by Oxford University Press Inc., New York

© Melissa M. Terras 2006

The moral rights of the author have been asserted
Database right Oxford University Press (maker)

First published 2006

All rights reserved. No part of this publication may be reproduced,
stored in a retrieval system, or transmitted, in any form or by any means,
without the prior permission in writing of Oxford University Press,
or as expressly permitted by law, or under terms agreed with the appropriate
reprographics rights organization. Enquiries concerning reproduction
outside the scope of the above should be sent to the Rights Department,
Oxford University Press, at the address above

You must not circulate this book in any other binding or cover
and you must impose the same condition on any acquirer

British Library Cataloguing in Publication Data
Data available

Library of Congress Cataloguing in Publication Data
Terras, Melissa M.

Image to interpretation : an intelligent system to aid historians in
reading the Vindolanda texts / Melissa M. Terras; with a foreword by
Alan K. Bowman and Michael Brady.

p. cm.

"Chapters 4, 5, and Appendix A coauthored with Dr. Paul Robertson."

ISBN-13: 978-0-19-920455-7 (alk. paper)

ISBN-10: 0-19-920455-1 (alk. paper)

1. Vindolanda tablets. 2. Vindolanda Site (Chesterholme, England)
3. Inscriptions, Latin—England—Chesterholme. 4. Wooden tablets—
England—Chesterholme. 5. Chesterholme (England)—Antiquities, Roman.
6. Paleography—Data processing. 7. Artificial intelligence—
Research. 8. Connectionism. 9. Minimum description length
(Information theory) 10. Image analysis. I. Robertson, Paul, 1956-II. Title.
DA147.V56.T47 2006

936.2'881—dc22

2006021796

Typeset by SPI Publisher Services, Pondicherry, India

Printed in Great Britain
on acid-free paper by
Biddles Ltd., King's Lynn

ISBN 0-19-920455-1 978-0-19-920455-7

1 3 5 7 9 10 8 6 4 2

Acknowledgements

This monograph stems from my doctoral thesis (Terras 2002), and I am indebted to my academic supervisors for their support and guidance: Professor Mike Brady (Department of Engineering Science, University of Oxford), and Professor Alan Bowman (Centre for the Study of Ancient Documents (CSAD), University of Oxford). Much of this work would not have been possible without the input, advice, and generosity of Dr Paul Robertson, formerly of the Robots Research Group at Oxford, and now Research Scientist of the Computer Science and Artificial Intelligence Lab (CSAIL) at Massachusetts Institute of Technology.

The research in Chapter 2 would have been impossible without the consent and enthusiasm of (in no particular order): Professor Alan Bowman, Dr Roger Tomlin (Wolfson College, University of Oxford), and Professor David Thomas (University of Durham).

Due to the interdisciplinary nature of this work, there are a number of other scholars to whom I am indebted: Dr Charles Crowther (CSAD), Dr John Pearce (formerly of CSAD, now of the Classics Department, King's College London), Dr Xiabo Pan, Dr Veit Schenk, and Dr Timor Kadir (all formerly of the Robots Research Group, Engineering Science, Oxford), Professor Seamus Ross (Humanities Advanced Technology and Information Institute, University of Glasgow), Edward Vanhoutte (Centrum voor Teksteditie en Bronnenstudie, Koninklijke Academie voor Nederlandse Taal-en Letterkunde, Ghent), and Professor Susan Hockey (School of Library, Archive, and Information Studies, University College London). My doctoral thesis was examined by Dr Marilyn Deegan (Centre for Computing and the Humanities, King's College London), and Professor Hilary Buxton (Department of Informatics, University of Sussex), and I thank them for their constructive criticism.

Thanks also to my friends: Natalie Walker, Dr Fleur Taylor, Dr Alison Kay, Thom Falls, Rosalind Porter, Dr Verity Platt, Claire Easingwood, David Beaumont, Susan Goldie, Dr Duncan Robertson,

and Dr Justin Pniower; my parents: Margaret and Robert Terras; and my grandparents: Helen and James Nelson, for their interest and support. Final thanks to Os, for his support during the writing of this monograph.

Foreword

This monograph presents an unusual and highly original piece of research. Its origin lay in our conviction, developed in the 1990s, that information technology offers the possibility of developments in techniques of image-capture and signal-processing that would help palaeographers and historians to read damaged, abraded, or otherwise illegible ancient texts. It seemed to us that approaches to the complex scientific and cognitive problems involved needed to find some middle ground between thinking either that technical solutions to all such problems exist and it is simply a case of finding the right set of existing standard procedures, or that it is impossible for scientific procedures to address the problems of really difficult material because the palaeographer/historian is him- or herself unsure what he or she wants to see and to interpret.

In approaching this subject from a background of English literature and art history, allied to a training in computing science and humanities computing, Dr Terras strongly emphasizes that the application of signal-processing techniques in reading, restoring, and interpreting damaged documents (in this case the Roman writing-tablets from Vindolanda, near Hadrian's Wall) is not and cannot possibly be a matter of developing some 'automated' procedure that will 'read' the texts for the historian. She begins from the principle that it is essential to understand the cognitive and intellectual processes involved in the decipherment of ancient texts and the ways in which the expert deploys his knowledge. This is a recursive procedure by which he gradually augments his reading and his understanding of the text, based on visual perception of an image which may itself be the product of a sophisticated process of digital image-capture, that is, one which offers a better image than the unaided human eye can achieve.

Of course, we lack a comprehensive theory of human visual perception or, indeed, of human cognition. In order to build computational systems that can aid, though not replace, the historian, it is first necessary to take advantage of what is known about

computational modelling of perception. Crudely put, this has two complementary aspects: the treatment of the array of intensities (brightnesses) that comprise the image, in a way that parallels the extraction of 'features' in the first stages of human vision, and the identification, subsequently the mobilization, of knowledge of what the image 'means'. In approaching the first of these two aspects, Dr Terras draws upon state-of-the-art understanding of signal processing, which were being developed by colleagues in engineering science as she conducted her research. Similarly, in approaching the second aspect, Dr Terras has adopted a number of ideas from perceptual psychology and artificial intelligence, in particular the theory of reading developed by Rumelhart and McClelland and the introspective agent (software) architecture developed by her fellow researcher Dr Paul Robertson. The result is both a novel synthesis of freshly developed ideas and a major test of the competence of those ideas.

The product is a monograph which breaks new ground in a genuinely interdisciplinary manner and will be of great interest to palaeographers and historians in offering exciting possibilities for enhancing the legibility and understanding of damaged ancient documents. And, as an extremely challenging application of ideas from image analysis and artificial intelligence, it will be of considerable interest to engineers and computer scientists.

A. K. Bowman
J. M. Brady

Contents

<i>List of Tables</i>	xii
1. Introduction	1
1.1 Vindolanda	2
1.2 The Vindolanda Texts	5
1.3 Image-Processing Techniques	9
1.4 An Interface for the System	12
1.5 Papyrology and Computing	12
1.6 Research Approach	17
1.7 Chapter Overview	22
I. Knowledge Elicitation and the Papyrologist	25
2. How do Papyrologists Read Ancient Texts?	27
2.1 Papyrology Discussed	28
2.2 Knowledge Elicitation: A Brief Guide	38
2.3 Knowledge Elicitation and Vindolanda	41
2.4 Associated Techniques: Content and Textual Analysis	47
2.5 Results	51
2.6 General Observations	72
2.7 Preliminary Conclusions	75
2.8 Models of Reading and Papyrology	77
2.9 Conclusion	82
3. The Palaeography of Vindolanda	84
3.1 Palaeography	85
3.2 Knowledge Elicitation and Palaeography	91
3.3 Information Used When Discussing Letter Forms	93
3.4 Derived Encoding Scheme	98
3.5 Building the Data Set	101
3.6 The Use of the GRAVA Annotator	105
3.7 Annotating the Characters	107
3.8 Results	110

3.9	Gathering Corpus Textual Data	111
3.10	How Representative is the Corpus?	113
3.11	Letter Forms	115
3.12	Conclusion	118
II.	Image to Interpretation	121
4.	Using Artificial Intelligence to Read the Vindolanda Texts (co-authored with Dr Paul Robertson)	123
4.1	Introduction to the GRAVA System	124
4.2	The Original System Demonstrated	126
4.3	Preliminary Experiments	129
4.4	The Construction of Character Models	132
4.5	The New System Explained	133
4.6	An Agent In Action	135
4.7	Conclusion	137
5.	Results (co-authored with Dr Paul Robertson)	138
5.1	Using Ink Data for Ink Tablets	139
5.2	Using Ink Models for an Unknown Phrase	143
5.3	Using Ink Models for Stylus Tablets	145
5.4	Using Stylus Models for Stylus Tablets	147
5.5	Analysing Automated Data	148
5.6	Future Work	149
5.7	Conclusion	151
6.	Conclusion	153
6.1	Knowledge Elicitation	154
6.2	Expanding the Existing System	158
6.3	Application Development	163
6.4	Evaluation	165
6.5	The Wider Scope	168
6.6	In Conclusion	169
<i>Appendix A. Using a Stochastic MDL Architecture to Read the Vindolanda Texts</i> (co-authored with Dr Paul Robertson)		171

Contents

xi

<i>Appendix B. Annotation</i>	200
<i>Appendix C. Vindolanda Letter Forms Corpus</i>	211
<i>References</i>	231
<i>Index</i>	249

List of Tables

2.1. The encoding scheme resolved from an analysis of the Vindolanda textual data.	49
2.2. Comparison of different individuals from transcripts of discussions.	52
2.3. Excerpt from Expert B's discussion of III 663 illustrating the appropriation of reading levels to the text, and demonstrating the flow of the discussion between different levels.	58
2.4. Likelihood of one reading level following another sequentially in the discussions.	60
2.5. Relationship of main and subsidiary topics within the Vindolanda apparatus.	62
2.6. Key words in the discussions regarding stylus tablets.	67
2.7. Features of the reading levels identified in the discussions regarding the ink and stylus texts.	75
B.3. Region identifier codes.	206

1

Introduction

We can try a little experiment. Let us resort to the fiction of programming an information transducer, a machine to read [ancient] texts. While so far only human beings have learned it, it is equally possible, and may one day be tried, to teach this skill to a machine . . .

(Reiner 1973: 6)

The ink and stylus texts discovered at Vindolanda are a unique resource for scholars of the Roman occupation of Britain. This book details the development of an appropriate knowledge-based computational system to aid papyrologists in the reading of the stylus texts. This system uses procedural information elicited from the experts who require such a system, and input from image-processing techniques, to output plausible interpretations of the texts.

The texts from Vindolanda (modern Chesterholm), a Roman fort on the Stanegate near Hadrian's Wall, are an unparalleled source of information regarding the Roman army for historians, linguists, palaeographers, and archaeologists. The visibility and legibility of the hand-writing on the ink texts can be improved through the use of infrared photography. However, due to their physical state, the stylus tablets (one of the forms of official documentation of the Roman army) have proved almost impossible to read. This chapter provides an introduction to Vindolanda and the ink and stylus texts discovered there, before discussing the developments in image processing that have allowed some features of the stylus tablets to be detected. The focus of this book is then established: the attempt to construct a computer system to aid the papyrologists in the reading

of the stylus texts. Background information is given regarding the use of computing in the field of papyrology, and the use of minimum description length (MDL) as a means to compare and contrast complex information within the field of image processing. An MDL-based system, GRAVA, is introduced that allows the integration of image and textual data into a reasoning system to generate possible interpretations of the text contained within images of the Vindolanda tablets. Finally, an overview of each chapter of the monograph is provided.

1.1. VINDOLANDA

The Roman fort of Vindolanda,¹ situated in the middle of the Tyne–Solway isthmus a mile south of Hadrian’s Wall, has provided historians and archaeologists with rich information regarding the Roman occupation of Britain, and life in one community on the outermost edge of one of Rome’s most remote provinces. Occupied by a garrison of the Roman army from the late first century AD until the abandonment of the northern British frontier three centuries later, Vindolanda was part of a network of forts along the Stanegate, a road and ditch system in development following the governorship of Agricola (AD 78–84) which was the main line of the frontier system before the construction of Hadrian’s wall (122–5).² Vindolanda’s continuing military importance after the construction of the Wall is demonstrated by the building of two stone forts to replace the earlier turf and timber structures, the first perhaps built in the latter part of the second century AD, the second in the late third century.

¹ Vindolanda means ‘white lawns’ or ‘white fields’. For further information regarding the Roman fort at Vindolanda, see Breeze and Dobson (1976), Birley (1977), Bidwell (1985, 1997). Good introductions to the Vindolanda texts are provided by Bowman (1994, 2003). Additionally, Vindolanda Tablets Online, <http://vindolanda.csad.ox.ac.uk/>, provides much contextual and historical information about the fort as well as presenting the Vindolanda tablets themselves.

² Bidwell (1997) provides a general introduction to Roman forts in Britain, and Millet (1995) gives a general overview of Roman Britain.

The documents found at Vindolanda date to the pre-Hadrianic timber forts. The vast majority belong to the period 97–103, when the fort was occupied by various auxiliary units: the First Cohort of Tungrians,³ the Ninth Cohort of Batavians,⁴ and perhaps the Third Cohort of Batavians, and a detachment of cavalry from the Spanish *Ala Vardullorum* (Bowman and Thomas 1994: 24). The names on the Vindolanda tablets suggest the non-citizen recruits had mostly Celtic and Germanic origins: Gallia Belgica and Germany, as well as Dacia, Italy, Spain, and Greece. Native troops were generally not stationed within their province of origin.

The Stanegate, and later Hadrian's Wall, were permeable frontiers, designed to control the movements of the tribes within the border zone, protect the areas of northern England most recently taken under direct Roman rule, and to regulate commerce between Roman Britain and its neighbours. The role, size, and needs of the forts along this frontier meant they were integrated into common society in northern England: to the west of the Hadrianic fort at Vindolanda was a small civilian settlement, or *vicus*, which serviced the needs of the garrison. It is difficult to separate the military and civilian activities of the Roman army at Vindolanda, as it had much interaction with the local area, and the documents found there provide information regarding the wider context of life in Roman Britain, as well as the operations of the Roman army during that period.

Despite Vindolanda's northerly setting, it was far from isolated from the rest of the Roman Empire: roads, built by the Romans to support the conquest of Britain, ensured rapid communications with other garrisons along the Stanegate (and, later, Hadrian's Wall), the south (London could be reached in less than a week), and the Continent. The documents discovered at Vindolanda, from Britain and perhaps the Gallic region as well, provide historians of this period with a unique resource regarding life at the fort and its surroundings, giving an insight into the military units, personnel,

³ The strength report for the First Cohort of Tungrians (Tablet 154, Bowman and Thomas (1994: 90) indicates that its actual strength was 752, including six centurions.

⁴ There is no direct evidence for the strength or nature of this unit (Bowman and Thomas 1994: 23).

activities of the army, supplies, and foodstuff required by the fort, the social and legal aspects of the fort and its vicinity, and information about the language used within everyday Roman society.

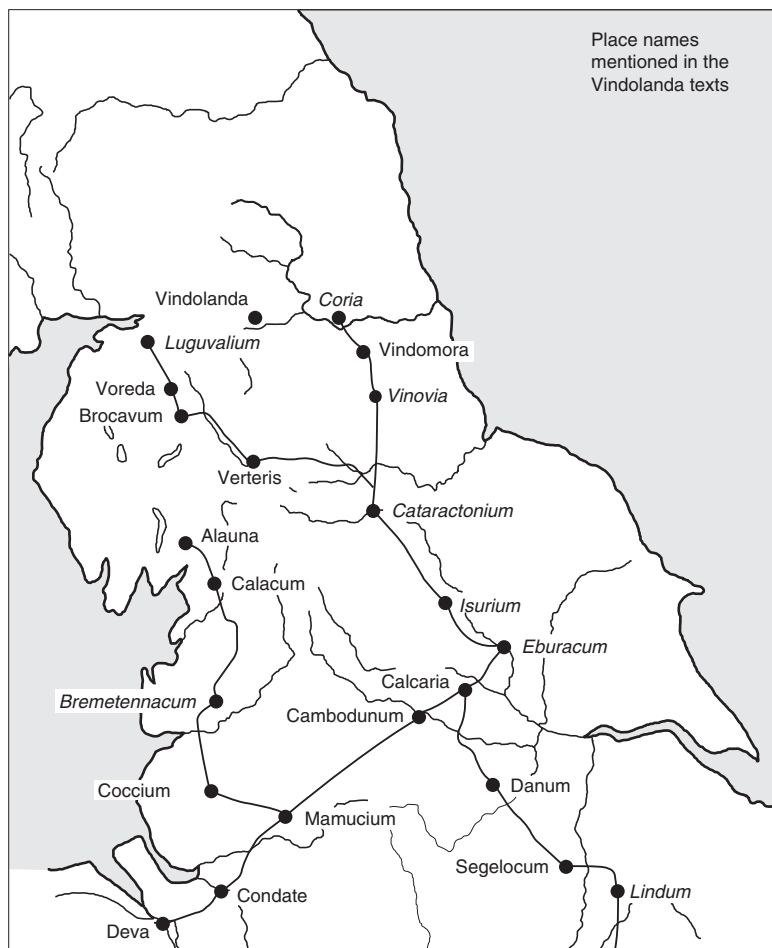


Figure 1.1. The Roman road network and settlements in northern Britain. Only the principal roads are shown. Place names which possibly occur in the Vindolanda texts are shown in *italic*. The number of place names mentioned in the Vindolanda tablets indicate the high level of communication between different Roman settlements. Image © Professor Alan Bowman, used with permission.

1.2. THE VINDOLANDA TEXTS

Textual sources for the period in British history from AD 90 to 120 are rare. Historians of this period have largely had to base their reconstructions of the Roman occupation of Britain on one historical account, the *Agricola* of Tacitus, a number of inscriptions, and archaeological sites and their artefacts.⁵ The ink and stylus tablets are a unique and extensive group of written documents from the Roman army in Britain, and provide a personal, immediate, detailed record of many aspects of the Roman fort at Vindolanda from AD 90 to 120 (Bowman and Thomas 1983, 1994, 2003; Bowman 1994, 1999, 2003). The texts are an unparalleled resource for classical historians as they ‘now cast a flood of light—or at least a galaxy of pinpoints of light—upon a Dark Age in northern Britain, upon the Roman army, its logistics, organization and social structure, and upon the spoken and written language of the time and milieu’ (Tomlin 1996: 463).

Preserved in the remains of the fort,⁶ the first ink tablets were unexpectedly discovered during archaeological excavations at Vindolanda in March 1973,

when a modern pipe trench outside the scheduled area was being widened... Amongst the planking and shaving, some exceptionally thin and regular slivers of wood were seen, and in one case, two slivers were stuck together. When casually prized apart, in the excavation trench, ink writing

⁵ Aside from the Vindolanda documents, the main sources of textual information regarding Roman Britain are narrative histories, inscriptions, and coins. Examples of these can be found in Ireland (1986: ch. 9). Further examples of inscribed objects found at Vindolanda, such as sherds, tile, and quern fragments, can be found in Tomlin (1985).

⁶ ‘The tablets were deposited in layers of bracken and straw flooring in successive occupation levels in the buildings and on the street outside. These deposits, some of which show signs of incineration, look like the dumping of rubbish and contain a wide variety of other organic remains and artefacts. The location of the early finds suggested that the presence of organic remains, in particular human excreta, might have been significant in creating chemical conditions crucial for the preservation of the wooden tablets. As the excavation progressed, however, tablets were found in areas where these conditions did not obtain, and it now seems likely that the damp, anaerobic environment is sufficient to account for the preservation’ (Bowman 2003: 14–15). For a discussion regarding the burning of out-of-date material at Vindolanda, see Birley (1997) and Bowman and Thomas (2003: 12).

was immediately evident, although at the time it could not be read, and within fifteen minutes it had become invisible to the naked eye... (Birley *et al.* 1993: 11)

Birley explains his exhilaration when it became apparent that the Vindolanda excavations were yielding Roman correspondence: 'If I have to spend the rest of my life working in dirty, wet trenches, I doubt whether I shall ever again experience the shock and excitement I felt at my first glimpse of ink hieroglyphs on tiny scraps of wood' (1977: 132).⁷ Careful and delicate analysis of the excavated soil has yielded many more ink texts since; the stylus tablets are much easier to identify, given their 'distinctive pine wood, grain, and shape' (Birley *et al.* 1993: 13). Conserved and housed at the British Museum,⁸ the ink and stylus tablets have been the focus of intense scholarly scrutiny for the last thirty years, providing much needed primary information regarding the language, society, and culture of Vindolanda, Roman Britain, and the Roman Empire. The archaeological site at Vindolanda continues to be excavated,⁹ with the hope of finding further writing tablets as well as significant archaeological evidence.

1.2.1. The Ink Tablets

The ink tablets, carbon ink written on thin¹⁰ leaves of wood cut from the sapwood of young trees, have proved the easiest to decipher. In most cases, the faded ink can be seen clearly against the wood surface by the use of infrared photography,¹¹ a technique used frequently

⁷ Further discussion regarding the discovery of the tablets can be found in Birley (1999).

⁸ An explanation of the cleaning, preservation, and photographic techniques used can be found in Birley (1977: 134) and Birley *et al.* (1993).

⁹ See <http://www.vindolanda.com/> for updates regarding current excavations.

¹⁰ The tablets are between 1 and 3 mm thick, indicating how fragile they are.

¹¹ Multi-spectral imaging involves illuminating an object with several different frequency bands: infrared, red, green, blue, and ultraviolet, which are then combined. Filters can be used to focus on information that is sensitive to irradiation by the non-visible portions of the light spectrum. Infrared (IR) photography is useful in imaging ancient texts as it uses wavelengths that are longer than those of visible light—typically 0.8 microns to 1.5 microns. IR is good for detecting black carbon-based

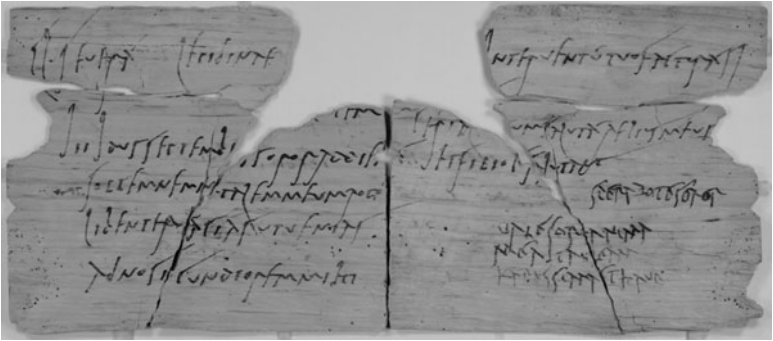


Figure 1.2. Ink text 291, captured by digital infrared photography. This diptych contains a letter to Lepidina, the wife of Flavius Cerialis who was the prefect of the Ninth Cohort of Batavians stationed at Vindolanda, from Claudia Severa, the wife of Aelius Brocchus. The letter is a warm invitation to Severa's birthday celebration. Whilst the body of the text is written by a scribe, the closing greeting, bottom right, is in Severa's own hand: this tablet almost certainly contains the earliest known example of writing in Latin done by a woman. See Bowman and Thomas (1994: 256) for a full transcription.

in deciphering ancient documents (Bearman and Spiro 1996; Zuckerman 1996; Knox *et al.* 1997; Bearman *et al.* 1998; Booras and Chabries 2001; Chabries and Booras 2001; Chabries *et al.* 2003). The majority of the several hundred writing tablets that have been transcribed so far contain personal correspondence, accounts and lists, and military documents (Bowman and Thomas 1983, 1994, 2003).

1.2.2. The Stylus Tablets

The two hundred stylus tablets found at Vindolanda conform to the format used by the Roman army and more generally throughout

ink because (a) IR corresponds to the electromagnetic wavelengths associated with absorption and emittance of heat; and (b) black ink is a stronger absorber or emitter than the surrounding vellum, papyrus, or wood. For example, in the (non-visible) Infra-Red spectrum, the black ink on a blackened papyrus scroll, such as those from Herculaneum and Petra (Chabries *et al.* 2003), or the faded carbon ink on the Vindolanda ink texts, can be clearly differentiated. For a description of how this technique was used to read the Vindolanda texts, see Birley *et al.* (1993: 103–6).



Figure 1.3. Stylus tablet 836, one of the most complete stylus tablets unearthed at Vindolanda. This text is a letter from Albanus to Bellus, containing a receipt and further demand for payment of transport costs (Ch. 2 gives a transcription of this text). The incisions on the surface can be seen to be complex, whilst the woodgrain, surface discoloration, warping, and cracking of the physical object demonstrate the difficulty papyrologists have in reading such texts.

the Empire (Turner 1968; Fink 1971; Lalou 1992; Renner 1992). It is suspected that their subject matter will differ to some extent from the ink tablets as similar finds indicate that stylus tablets tended to be used for documentation of a more permanent nature, such as legal papers, records of loans, marriages, contracts of work, the sale of slaves, etc. (Renner 1992). However, the palaeographic and linguistic characteristics of the stylus tablets may reasonably be expected to be similar to the ink texts as they are contemporaneous documents from the same sources.

Manufactured from softwood with a recessed central surface, the hollow panel of the stylus tablets was filled with coloured beeswax. Text was recorded by incising this wax with a metal stylus,¹² and

¹² For examples of stylus pens found during the archaeological excavations at Vindolanda, and a further discussion of the writing materials used, see Birley (1999: 17–27).

tablets could be reused by melting the wax to form a smooth surface. Unfortunately, in nearly all surviving stylus tablets¹³ the wax has perished, leaving a recessed surface showing the scratches made by the stylus as it penetrated the wax.¹⁴ In general, the small incisions are extremely difficult to decipher.¹⁵ Obtaining a reading is complicated further by the pronounced woodgrain of the wood (often fir or larch) used to make the stylus tablets, staining and discoloration of the surface, damage over the past two thousand years, and the palimpsest nature of the reused tablets. A skilled reader can take several weeks to transcribe one of the more legible tablets, and several months to transcribe a portion of some of the more damaged texts, whilst many have to date defied any interpretation. Prior to the research described here, the only way for the papyrologists to detect incisions in a tablet was to move the text around in a bright, low raking light. In doing this, indentations are highlighted and candidate writing strokes become apparent through the movement of shadows, although this proves to be a frustrating, time-consuming, and inefficient way of reading the texts.

1.3. IMAGE-PROCESSING TECHNIQUES

A digital image is a representation of a two-dimensional image as a finite set of digital values, called *picture elements* or pixels, with the

¹³ It is estimated that *c.* 2,000 such tablets exist outside Egypt (Renner 1992).

¹⁴ Only one stylus tablet, 836, has been found so far with its wax intact. Unfortunately this deteriorated during conservation, but a photographic record of the waxed tablet remains to compare the visible text with that on the reused tablet. A discussion regarding this tablet can be found in Bowman and Tomlin (2005: 10).

¹⁵ Even Roman readers sometimes had difficulty with reading stylus tablets. In a scene from *Pseudolus*, by the Roman playwright Plautus, the eponymous character, the slave Pseudolus, is unimpressed by the writing on a wax tablet sent to his master, Calidorus, by the young man's lover: 'PSEUDOLUS: All these letters, they seem to be playing at fathers and mothers, crawling all over each other. CALIDORUS: Oh, if you're going to make a joke of it—PSEUDOLUS: It would take a Sibyl to read this gibberish; no one else could make head nor tail of it. CALIDORUS: Why are you so unkind to those dear little letters, written on that dear little tablet by that dear little hand? PSEUDOLUS: A chicken's hand was it? A chicken surely scratched these marks...' (Plautus, *Pseudolus* 1. 1.23–30).

information stored in a large rectangular array within computer memory. Computers can be used to analyse these arrays of numbers carrying out a task called ‘image processing’. The computer will implement algorithms that compare adjacent numbers in the array, looking for large changes over short segments of the image or characterizing patterns that may be found in the numbers. Image-processing techniques can be used to detect or enhance information (see Gonzalez and Woods (1993) for an introduction).

In 1998 the Department of Engineering Science and the Centre for the Study of Ancient Documents at the University of Oxford were jointly awarded a research grant by the Engineering and Physical Sciences Research Council (EPSRC) to develop techniques for the detection, enhancement, and measurement of narrow, variable depth features inscribed on low contrast, textured surfaces (such as the Vindolanda stylus tablets). To date, the project has developed a wavelet filtering technique¹⁶ that enables the removal of woodgrain from images of the tablets to aid in their transcription (Bowman *et al.* 1997). In addition, a technique called ‘shadow stereo’ or ‘phase congruency’ has been developed, in which the camera position and the tablet are kept fixed; but a number of images are taken where the tablet is illuminated by a strongly orientated light source.

If the azimuthal direction of the light source (that is, the direction of the light source if the light were projected directly down on to the table) is held fixed, but the light is alternated between two elevations,

¹⁶ Filtering is a generic term that refers to methods for extracting limited information from a signal or image, or for removing certain kinds of detail – such as noise. It is used in the field of digital image processing to enhance digital image data (image sharpening, image smoothing, and familiar techniques such as red eye reduction), but more generally for image analysis, and pattern recognition. Wavelet filtering is an alternative to conventional digital filtering methods (such as the Fourier Transform). The wavelet transform has the advantage that a ‘feature’ can be localized simultaneously both in time (signal) or space (image) as well as in the frequency domain, hence giving a time-frequency representation of the signal. The time-domain signal is passed between various high-pass and low-pass filters, which filters out either high frequency or low frequency portions of the signal. This procedure is repeated, to provide information at a succession of temporal/spatial scales, and every time some portion of the signal corresponding to some frequencies is removed from the signal. In the case of the Vindolanda tablets, this technique was used to remove woodgrain in images of the text, and to extract the strokes of letters. See Bowman *et al.* (1997) for a discussion on the application of wavelet transforms to the Vindolanda texts, and Burrus *et al.* (1997) for a general introduction to wavelet transforms.

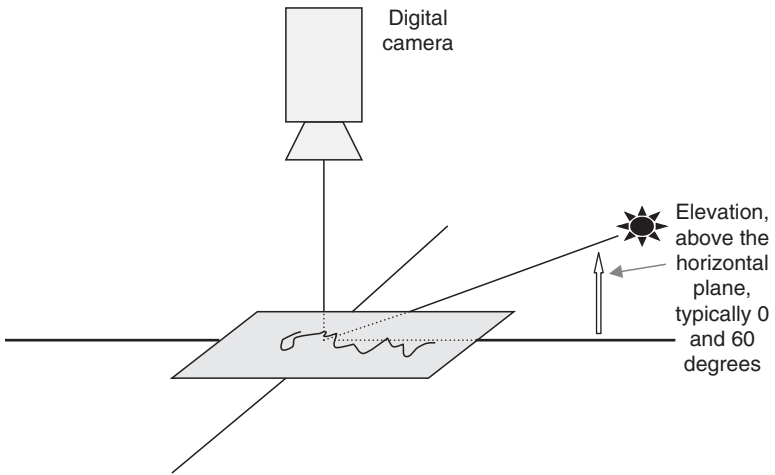


Figure 1.4. Shadow stereo in action. Whilst the elevation of the light source changes around the tablet, the camera and tablet remain stationary. Shadows cast by incisions move in a predictable manner, and image-processing algorithms can determine which incisions are most likely to be candidate writing strokes of characters, as opposed to noise.

the shadows cast by incisions will move but stains on the surface of the tablet remain fixed. This strongly resembles the technique used by some papyrologists who use low raking light to help them read the incisions on the tablet (Molton *et al.* 2003). Edge detection is accomplished by noting the movement of shadow to highlight transitions in two images of the same tablet, and so candidate incised strokes can be identified by finding shadows adjacent to highlights which move in the way that incised strokes would be expected to move (Schenk 2001; Schenk and Brady 2003). Although this is not a standard technique in image processing, encouraging results have been achieved so far (see S. 4.3.1), and a mathematical model has been developed to investigate which are the best angles to position the light sources (Molton *et al.* 2003). Work currently being undertaken is extending the performance and scope of the algorithms (Pan *et al.* 2004), and the papyrologists are beginning to trust the results and suggestions which are being made about possible incisions on the tablets (Bowman and Tomlin 2005). Future work will be

done in relating the parameters of analysis to the depth profile of the incisions to try to identify different overlapping writing on the more complex texts, and an automated annotation process is being developed to extract stroke information in a format that can be used as input for the system described in this monograph (Robertson *et al.*, forthcoming; see also S. 5.5).

1.4. AN INTERFACE FOR THE SYSTEM

Whilst the techniques discussed above have had some success in analysing the surfaces of the tablets, it was essential to develop a way of aiding the papyrologists in utilizing generated results. Although the developed algorithms could be easily added to readily available image manipulation software,¹⁷ allowing others to apply the algorithms themselves, it would do little to actually provide a tool that would actively help the papyrologists in the transcription of texts; a complex process which has been described as ‘teasing information out of material which is all too often barely legible, fragmentary, or obscure (or all three at once)’ (Bowman 1994: 10). The focus of this monograph is to investigate the possibility of developing an intelligent system that can aid the papyrologists in their task.

1.5. PAPYROLOGY AND COMPUTING

The use of computing in the field of papyrology¹⁸ has enjoyed some notable successes. There are many established imaging projects, such as those at the Centre for the Study of Ancient Documents,¹⁹ and the Oxyrhynchus Papyri Project;²⁰ excellent database

¹⁷ e.g. using the Visual C++ plugin with Adobe PhotoShop, see <http://partners.adobe.com/public/developer/photoshop/devcenter.html>.

¹⁸ Papyrology can simply be defined as obtaining ‘a body of knowledge... from the study of papyri’. It is now taken to cover ‘as a matter of convenience... the study of all materials carrying writing... done by a pen’ (Turner 1968: vi). See Ch. 2 for further discussions regarding ‘papyrology’.

¹⁹ <http://www.csad.ox.ac.uk>

²⁰ <http://www.csad.ox.ac.uk/POxy/>

projects and systems such as the Duke Data Bank of Documentary Papyri,²¹ and APIS (Advanced Papyrological Information System);²² and repositories of information in a user-friendly format such as the Perseus Digital Library.²³ Individual projects, such as Vindolanda Tablets Online²⁴ and Inscriptions of Aphrodisias,²⁵ present individual collections of texts combined with images, transcriptions, and editorial comment online. Many standards are already in place for the digitization²⁶ and markup²⁷ of ancient texts, and papyrologists are making more use of the kind of image manipulation tools provided by the likes of PhotoShop (Bagnall 1997).

Simple image-processing techniques have been used since the late 1970s to aid scholars in the reading of individual manuscripts (from the ancient, medieval, and modern periods),²⁸ such as

²¹ <http://odyssey.lib.duke.edu/papyrus/texts/DDBDP.html>

²² <http://www.columbia.edu/cu/lweb/projects/digital/apis/>

²³ <http://www.perseus.tufts.edu>

²⁴ <http://vindolanda.csad.ox.ac.uk/>

²⁵ <http://insaph.kcl.ac.uk/>

²⁶ See <http://www.csad.ox.ac.uk/CSAD/Images.html> for the aims and methods of digitizing ancient documents and inscriptions at the Centre for the Study of Ancient Documents, and the Technical Advisory Service for Images website for a good general introduction to digitisation: <http://www.tasi.ac.uk/>.

²⁷ EpiDoc (short for 'Epigraphic Documents') is a software- and hardware-independent digital publication and interchange specification for scholarly and educational editions of inscribed and incised texts in Greek, Latin, and other languages emanating from the Ancient Greek, Roman, and nearby civilizations. EpiDoc utilizes TEI, the application of XML (Extensible Markup Language) developed by the Text Encoding Initiative. EpiDoc can also be applied to other non-inscribed texts: the Vindolanda tablets have been marked up in EpiDoc to allow for easy textual manipulation, transmission, and interchange. See <http://www.ibiblio.org/telliott/epi-doc/> for the EpiDoc homepage, <http://www.tei-c.org/> for information about the TEI, <http://www.w3.org/XML/> for general information about XML, and <http://vindolanda.csad.ox.ac.uk/tablets/TVdigital-1.shtml> for information about how EpiDoc has been applied to the Vindolanda tablets.

²⁸ The techniques of Digital Image Processing, which resulted from developments in the National Aeronautics and Space Administration (NASA) space program, were applied to manuscript material as early as 1978, in order to use 'an electronic camera and the image-processing techniques developed for space photography' to decipher most of an erased ex-libris from a 14th-cent. manuscript: 'Contrast enhancement or optimization is the key technique; where fading has been a problem, this can include equalization of contrast in different areas of the image. Spatial filtering can be used to removed shading and permit even greater contrast enhancement. Other filtering techniques which 'detect' and enhance edges are useful in emphasizing the strokes of a written text. Additional processing can then be employed to differentiate between

the *Beowulf* manuscript²⁹ (Kiernan 1991; Prescott 1997), and the carbonized scrolls from Herculaneum and Petra³⁰ (Ware *et al.* 2000; Booras and Chabries 2001; Chabries and Booras 2001; Chabries *et al.* 2003). Image manipulation tools are being developed which enable experts to virtually manoeuvre texts and light sources (Woolley *et al.* 2001; Lundberg 2002). Librarians and archivists are beginning to enquire how document imaging analysis can aid them in reading, retrieving, manipulating, and storing document data (Jameson 2004). However, research done in the area of image processing of ancient documents often concentrates on the computational element, with little focus being given to the needs of the experts trying to read the documents (see Stark (1992), Seales *et al.* (2000), Brown and Seales (2001), and Rivero *et al.* (2003) for examples of projects involving the imaging of ancient texts with little or no input from palaeographers or papyrologists). There is often very little consultation with the experts for whom the tools are being developed, and scant understanding of their requirements; projects tend to develop tools that they *think* will be useful, rather than determining how the experts operate, what they actually need, and what they can and will use to aid them in their task.

actual character strokes and random background noise, erasing much of the latter'. (Benton *et al.* 1979: 47).

²⁹ Image-processing algorithms were used with digital images of the *Beowulf* manuscript to improve radically the contrast, e.g. by stretching the range of brightness throughout the image, emphasizing edges, and suppressing random background noise (coming from the equipment rather than the document). This enabled areas of the text to become clearer, although it ultimately reinforced the thesis that it may be impossible to obtain a complete, unattested reading of the text: 'The value of digital image processing in the specific case of this manuscript is not that it suddenly reveals things not already visible in one way or another to palaeographers, but that it markedly improves the legibility of facsimiles for everyone else. Rather than settling the text, however, digital image processing may ultimately oblige us to admit that we do not and probably cannot have an established text of *Beowulf*' (Kiernan 1991: 26).

³⁰ The use of Multi-Spectral Imaging for the analysis of carbonized ancient documents has been pioneered by the Center for the Preservation of Ancient Religious Texts at Brigham Young University, Utah (<http://cpart.byu.edu/index.php>). Digital Imaging is used alongside Multi-Spectral Imaging analysis to analyse the generated images further, enabling some of the texts, which are illegible to the human eye, to be read.

No systems currently exist to support papyrologists in the *process* of reading ancient texts.³¹ Indeed, there is little research published which discusses how information is actually extracted from these texts (see S. 2.1.1) and there does not exist detailed cognitive and/or perceptual information-processing models of the papyrology process. From a cognitive psychology stance, although there has been much consideration of the processes involved in reading (see S. 2.1.2), few conclusions have been drawn as to how a reader would approach such damaged, fragmentary, foreign-language texts and construct a logical, acceptable meaning. Also, although image processing is an expanding field in the discipline of engineering science (see Gonzalez and Woods (1993) for an introduction) little work has been done on the role of knowledge and reasoning in the analysis and understanding of complex images. Proposals for integrating image analysis algorithms with techniques for the representation and mobilization of knowledge (the subject of the field of artificial intelligence) remain few.

³¹ Levison, and independently Ogden, considered the potential for the use of computers in the reconstruction of ancient manuscripts in the late 1960s but were hampered by lack of computing power (Levison 1965, 1967; Ogden 1969; Levison 1999). Wacholder and Abegg (1991) used a computer in an effort to reconstruct the unpublished Dead Sea Scrolls, and Finney (1999) uses computational multivariate analysis to document variation in New Testament Greek texts, but these are much simpler tasks than the one that is described here, due to the fact that they involve interrogation of textual data, rather than image data. Kleve *et al.* (1987) discuss the potential for the use of digital imaging in the reading of the Herculaneum texts, using graphics to trace letter forms to aid in comparison and identification, and suggesting the role that more advanced 'computer-enhanced digitized imaging' (p. 116) may play in future systems. However, the system described is one of recording, transposition, and manual comparison, and does not aim to aid the scholar in resolving ambiguity through any kind of computational intelligence. Software called SPI (System for Palaeographical Inspection), implemented at the Computer Science department of the University of Pisa (Aiolfi 1999; Aiolfi *et al.* 1999), was developed as means to support palaeographical observation and interpretation. A relational database works as the core of the system, where bitmaps of page images, of segmented characters, and of generated models of letters can be stored, recalled, and compared. This system was used to analyse a selection of codices held at the public library (Biblioteca degli Intronati) of Siena in Italy to test the real support that the software can offer to a traditional palaeographical analysis (Ciula 2004), although again the system relies on manual manipulation and sorting of data, and requires further development of the user interface and input from palaeographers to become a useful tool. Additionally, it focuses on the palaeography of letters, rather than process of identification and reading.

There were some major issues that needed to be tackled to develop a system to aid papyrologists in their task:

1. An understanding of how papyrologists operate was required.
2. Once this was achieved, knowledge elicitation needed to be carried out on all semantic levels of the process: from gathering the information used regarding character form and identification (palaeography), to an understanding of the type of language that was used in the texts.
3. A way of representing the elicited knowledge was required. A system was needed to capture and mobilize information regarding letter forms. Statistics regarding the use of language at Vindolanda had to be obtained.
4. It was necessary to find a way of linking this data to the existing image-processing techniques that have been developed to aid in the reading of the stylus tablets, so that stroke data from the feature detection algorithms could be compared with the information captured regarding character forms.
5. A way of comparing graphical and linguistic information had to be found, in order to generate probable interpretations of the texts.
6. A graphical user interface would have to be developed to deliver the system as a stand-alone application to the experts to aid them in their day-to-day task.

A major aim of this monograph is to contribute to the understanding of how papyrologists carry out their task. The type of information used by the experts regarding letter forms is also made explicit, and a way of representing and utilizing this information is described. Statistical knowledge about the language contained within the Vindolanda texts is linked to the work done on image processing through the use of a stochastic intelligent system which generates possible interpretations of image data. This demonstrates a proof of concept: that eliciting knowledge from experts can lead to the development of a computational system which replicates their behaviour, and that the gathering of this information can, in itself, give new insights into the behaviour of experts, and provide a tool to aid the experts in carrying out their task.

The development of this prototype system also contributes to engineering science and artificial intelligence, as it was accomplished

by developing a novel architecture based on minimum description length (see SS. 1.6.3, 4.1, and Appendix A), demonstrating that it is a viable technique for comparing and contrasting information across semantic levels. Some suggestions are made as to how this system could be delivered to the papyrologists as a desktop application, and work continues on the development of this system. This monograph tells the story of development to implementation, presenting the inter-disciplinary techniques used to construct a prototype system which can read in image data of the Vindolanda texts, and output plausible interpretations within a reasonable time frame.

1.6. RESEARCH APPROACH

1.6.1. Knowledge Elicitation

In order to identify the tools that could be built to aid the papyrologists in their transcription of the Vindolanda tablets, it was first necessary to try and gain an understanding of what the papyrology process actually entails. A programme of ‘knowledge elicitation’³² was undertaken (see S. 2.2), and from this, a model of how papyrologists approach and start to understand ancient texts was developed (see S. 2.8.3). This was then related to current cognitive psychology theories regarding the resolution of ambiguity in texts during the process of reading. There are many papers regarding this phenomenally complex process, but only a few computer programs have been implemented to test the models postulated. Parallels between one of these major theories, the ‘Interactive Activation and Competition Model of Word Recognition’, which aims to explain the word superiority effect, developed by McClelland and Rumelhart (1981, see S. 2.8.2), and the findings of the knowledge elicitation experiments were obvious. The model of how the papyrologists operate was taken as the basis for developing the architecture of the system described in Chapters 4 and Appendix A of this monograph.

³² A structured process used to acquire knowledge from human experts and learn from that data. The history, development, and techniques used in knowledge elicitation are covered in section 2.2.

1.6.2. Computer Aided Handwriting Recognition

The field of automated handwriting recognition is expansive and complex (see Impedovo (1993) for an introduction), but although there has been a great deal of work on pattern recognition algorithms aimed at recognizing text, the vast majority are irrelevant to this research. Most techniques are aimed at recognizing printed text, making them incompatible with the handwritten, cursive text found on the Vindolanda tablets. Work done on handwritten non-cursive text primarily approaches the problem with pattern class or neural network³³ learning; where the system learns individual characters from thousands of examples. There are simply insufficient examples to be able to train such a neural net, even from the corpus of Vindolanda ink texts. Other work, largely from the late 1960s and early 1970s, emphasized ‘syntactic pattern classification’: the idea that a character is composed of strokes that have a certain relationships to each other (see Connell and Brady’s approach to shape representation (1987), and Fu and Swain’s introduction to syntactic pattern recognition (1969)). The attempts to teach a machine to “read” text in this manner were hampered by the problem of stroke detection: image processing techniques were not developed enough to provide the necessary data. Due to the subsequent advances made regarding feature detection (see S. 4.3.1), a similar, syntactic way of modelling character information was adopted in this book (see S. 3.3), as a means of capturing a set of data with which to compare information extracted from the images of the text.

There have been previous attempts to use connectionist³⁴ inspired models of human reading as the basis on which to build systems to “read” cursive handwriting (Dodel and Shinghal 1995; Parisse 1996; Côté *et al.* 1998). These systems have had limited success: they are dependent on over-simplification of word shape and contour, the

³³ A particular model of a multiprocessor computer system used in perceptual psychology and in artificial intelligence, and based on the parallel organization of human information processing. This involves a network of relatively simple processing elements, like the neurons in the human brain, where the global behaviour is determined by the connections between the processing elements and element parameters. See Appendix A. no. 1.

³⁴ An approach in the field of artificial intelligence and cognitive science which models mental or behavioural phenomena with neural networks. See Appendix A no. 1.

models rely on strong contextual constraints, and they are only successful with small lexicons of 25–35 words. Appropriating these systems in an attempt to build a tool for the stylus tablets would be unsuccessful for the same reason that existing image-processing techniques could not be used to analyse the surface data: the data is too fragmentary, too noisy, and too complex. It has been suggested that using minimum description length may provide a way to successfully model systems when only sparse data exists (Robertson 2001), and this is the technique adopted here.

1.6.3. Minimum Description Length

A major puzzle in perceptual psychology has been how the brain reconciles different sorts of information, for example, colour, motion, boundaries of regions, or textures, to yield a percept (Eysenck and Keane 1997). For example, in image segmentation, by changing the texture model it is possible to increase or decrease the number of regions that are identified. Similarly, the number of boundary shapes that are found can be increased or decreased by changing the search parameters. The reading of ancient documents is just one (complex) example of an interpretation problem, where the individual is faced with visual ambiguity and competing information, which has to be reconciled and resolved with other types of information (in this case linguistic data) to generate a plausible solution. There has been substantial consideration, by psychologists and image-processing experts, of how these different processes are combined. One suggestion is that there is a common value that can be used to calculate the ‘least cost’ solution when comparing different types of information. This has been adopted by the field of artificial intelligence, in the concept of minimum description length (MDL).

First introduced in the late 1960s, and developed in the 1970s (Wallace and Boulton 1968; Rissanen 1978), MDL applies the intuition that the simplest theory which explains the data is the best theory.³⁵ MDL can be used as a means of comparing data in coding

³⁵ The basic concept behind MDL is an operational form of Occam’s razor, ‘Pluralitas non est ponenda sine necessitate’ or ‘plurality should not be posited without necessity’. Formulated by the medieval English philosopher and Franciscan

theory, in which the goal is to communicate a given message through a given communication channel in the least time or with the least power. MDL is a very powerful and general approach which can be applied to any inductive learning task, and can be used as a criterion for comparing competing theories about, or inductive inferences from, a given body of data. It is well suited to interpretation problems: where solutions can be generated by comparing unknown data to a series of models, when the most likely fit can generate plausible solutions to the problem.

MDL has been applied to numerous image-processing problems, to provide a means to compare information and choose between competing solutions. Leclerc used MDL in an attempt to solve the image partitioning problem, to delineate regions in an image that correspond to semantic entities in a scene (1989). Zhu and Yuille utilized MDL to develop a novel statistical and variational approach to image segmentation through the use of region competition (1996). Other recent research which uses MDL as a means of comparing and contrasting possible solutions in the field of image processing includes Lanterman *et al.* (1998), Rissanen (1999), and Gao and Ming (2000).

In his doctoral thesis ‘A Self Adaptive Architecture for Image Understanding’ Paul Robertson expands the remit of the type of information exchanged within an MDL-based stochastic systems architecture (2001). Although previous (known) research had limited the use of MDL to image understanding, Robertson incorporated linguistic information into his system to build an application that could effectively ‘read’ a handwritten phrase (see S. 4.1). Robertson’s GRAVA system uses a model of reading similar to the ‘Interactive Activation and Competition Model for Word Recognition’ (McClelland and Rumelhart 1981; see S. 2.8.2) combined with MDL (see S. A.2.1) and Monte Carlo³⁶

monk William of Ockham (c.1285–1349), the phrase has been adopted by communication theorists to suggest that one should not increase, beyond what is necessary, the number of entities required to explain anything (Forster 1999). The shorter an explanation, the better. Concise is good.

³⁶ Monte Carlo techniques are statistical methods for sampling using random numbers (or more often pseudo-random numbers), as opposed to deterministic algorithms. Monte Carlo methods use randomness and employ repetition to find an approximation to a solution, and are often used to find solutions to complex

methods (S. A.2.2) to propagate possible interpretations of a written text. Robertson's example is limited to the reading of a small phrase from a nursery rhyme. It was suggested in his thesis that this architecture could be suitable for the implementation of a system that deals with much more complex data. There was much work to be done to adapt Robertson's system to the needs of this research.

1.6.4. System Construction

In the research detailed here, the GRAVA system constructed by Robertson is adopted as the means by which to construct a system that can effectively 'read' and 'reason' about the texts found at Vindolanda. The original GRAVA system was significantly adapted to carry out this more complex task, and much intricate data needed to be collected in such a way that it could be mobilized efficiently, in order to make the adoption of GRAVA feasible.

First, an investigation was undertaken into the type of letter forms found at Vindolanda, which enabled an encoding scheme to be developed. Images of ink and stylus texts were then annotated, using this encoding scheme, to build up a corpus of letter forms which could be used to train the system (see Chapter 3). Statistics regarding the language used at Vindolanda were generated from the ink tablet corpus and prepared for integration with the system. The GRAVA system was then adapted, trained on the image corpus, and used to generate interpretations of the Vindolanda texts (see Chapters 4 and 5, and Appendix A). The resulting system was also tested with data that was generated from the automatic feature extraction algorithms by other members of the team (see Chapter 5 and Appendix A). It is demonstrated that such a stochastic MDL architecture provides a means by which to interpret images of the Vindolanda documents, and as such provides the basis for a stand-alone application with which to aid the papyrologists in their reading of the stylus texts.

mathematical problems that cannot easily be solved by other numerical methods. The name stems from the city's association with gambling, and the random sequence of numbers produced by the throw of dice or spin of a roulette wheel. See Hammersley and Handscomb (1964) for an introduction.

It should be stressed here that the construction of this system is not an attempt to build an ‘expert system’ that will automatically ‘read’ and provide a fixed transcription of the texts, thus negating the input of the papyrologists. Rather it is an attempt to build a system with which papyrologists will be able to mobilize disparate knowledge structures, such as linguistic and visual clues, and use these in the prediction process to aid in the resolution of the ambiguity of the texts. The system as it stands propagates possible solutions to the input data. Integration of this into an application which allows the papyrologist to adjust parameters of the system would enable them to maintain an explicit record of the alternative hypotheses developed as they attempt to read such a text, whilst suggesting possible solutions to aid them in their task. This would allow them to switch effortlessly between initially competing hypotheses, allowing them to see the development of their reading of the texts and trace any conclusions back to their initial thought processes—something which is difficult to do at present.

The overall aim in this research is to aid in the transcription of the stylus tablets, but it is hoped that the system may be eventually be used by papyrologists working on other texts (once a stand-alone system is made available). Additionally, this research has also provided a great deal of information regarding how papyrologists operate. The techniques used would be easily adaptable to other fields: there are many applications for a computer system that can effectively reason about data contained within images. Most importantly, it demonstrates that MDL can be used effectively as a unifying method to compare and resolve different forms of complex data, and that MDL can provide the basis for a cognitive visual system.

1.7. CHAPTER OVERVIEW

This monograph comprises of six chapters and three appendices, which cover the following topics:

- **Chapter 1** contains background and contextual information regarding the research, plus an overview of the focus of the monograph.

- **Chapter 2** asks the question: how do experts read ancient texts? After considering all available research on this topic, this chapter describes an investigation, using knowledge elicitation techniques, into how the experts read both the Vindolanda ink and stylus texts. The findings are rationalized into a connectionist model on which to base the development of the computer system detailed in Chapter 4 and Appendix A.
- **Chapter 3** concerns the letter forms used within the texts. It is shown that the stylus tablets should contain similar letter forms to the ink texts. An encoding scheme is developed to capture information about the letter forms by collating information discussed by the experts as they identify individual characters. A corpus of annotated images is built up to provide a data set on which to train the system developed in Chapter 4 and Appendix A.
- **Chapter 4** provides an overview of the architecture adopted in order to build a system to effectively ‘read’ the stylus texts, and generate possible interpretations of the text contained within the images. The background and architecture of the original system are described, and it is then explained how and why this system had to be adapted to deal with the complexity of the Vindolanda texts, before discussing the basic underlying structure of the system. More technical information about the system can be found in Appendix A.
- **Chapter 5** presents results, running the system on various sets of test data. It is shown that an MDL-based image architecture can take as input annotated images, and output plausible interpretations of these images in a reasonable time frame. As such, it is demonstrated that the GRAVA architecture provides the basis for a cognitive imaging system.
- **Chapter 6** details future work, suggesting how the developed system could be expanded, and eventually delivered to the experts as a stand-alone application to aid them in their task of reading the stylus tablets. The findings of the research are summarized, and its success evaluated.
- **Appendix A** describes in technical detail the implementation of the MDL-based GRAVA architecture used to read the stylus texts. The concepts of minimum description length and Monte Carlo methods, which make GRAVA so novel and attractive for this

purpose, are explored in depth. The objects that constitute the GRAVA system are detailed, and the Monte Carlo algorithm applied to agent selection is expounded. The development of information models, which the system utilizes to propagate realistic interpretations, is described.

- **Appendix B** details the encoding scheme used to annotate the images, and an example is given of an encoded character. An introduction to the corpus of annotated images is provided.
- **Appendix C** contains a graphical representation of the stroke data of all the individual characters contained in the annotated corpus, providing a resource for palaeographers and papyrologists.

I

Knowledge Elicitation and the Papyrologist

This page intentionally left blank

How do Papyrologists Read Ancient Texts?

It seems to me that translating from one language into another...is like looking at Flemish tapestries from the wrong side; for although the figures are visible, they are covered by threads that obscure them, and cannot be seen with the smoothness and color of the right side...And I do not wish to infer from this that the practice of translating is not deserving of praise, because a man might engage in worse things that bring him even less benefit.

(De Cervantes, *Don Quixote*, Second Part, Chapter 62 (2004: 873))

Before designing and building any tools to aid papyrologists in the reading of texts, it is a necessary requirement first to ask: just what does a papyrologist do when trying to read and understand an ancient text? A review was undertaken of all relevant research in this area, before an investigation was carried out to elucidate this process. This chapter provides an overview of other research that has questioned how papyrologists may read ancient documents, discusses the methodology used with the experts who read the Vindolanda texts to ascertain their thought processes when approaching a text, and presents, as a result of the investigation, a model of how papyrologists read ancient documents.

Researchers in the field of knowledge elicitation have developed many techniques and tools to interact with domain experts, with the aim to gather, resolve, and apply their knowledge, mainly to aid in the construction of computer systems. Knowledge elicitation techniques, such as structured interviews and think aloud protocols, were used with experts working on the Vindolanda texts to gather

quantitative and qualitative information about how papyrologists work. This resulted in an in-depth understanding of the ways different experts approach and reason about damaged and abraded ancient texts.

The process of reading, and making sense of, an ancient text is resolved into defined units, with characteristics about each being documented. General procedural information about the ‘papyrology process’ is also presented. Particular issues regarding problems in reading the Vindolanda stylus texts are highlighted, indicating areas in which computational tools may be able to aid the papyrologists in reading such texts. This results in a proposed model of how experts read ancient documents which is used in subsequent chapters as a basis for the development of a computer system.

2.1. PAPYROLOGY DISCUSSED

The term ‘papyrology’ emerged towards the end of the nineteenth century (*OED*, 2005). Although research had been carried out on papyri and associated documents in a varied if disorderly manner for many years before, the rapid expansion of available texts from discoveries made at archaeological sites, and the associated research community interested in these texts, gave rise to the need for a scientific term and systematization of the discipline. B. P. Grenfell, a pioneer of papyrology at Oxford University, explained the developing meaning of the term:

Papyrology has come to mean practically the study of Greek Papyri, including various substitutes for papyrus as writing-material, such as ostraca (bits of broken pottery), wooden or wax tablets, and after the second century vellum. Like epigraphy, papyrology is an aid to the study of Greek and Roman antiquity in its various departments, and is not an independent branch of enquiry. (1921: 1–2)

The readings generated from ancient documents provide one of the major primary information sources for classicists, linguists, archaeologists, historians, palaeographers, and scholars from associated disciplines. Although there has been some discussion of the history

of papyrology (Hunt 1922, 1930; Turner 1968; Pattie and Turner 1974; Gallo 1986; Pestman 1990; van Minnen 1995), and the contribution the transcription of such texts has made to both literary and non-literary classical studies (Hall 1913; Hunt 1914; Kenyon 1918; Winter 1936; Turner 1968; Ullman 1969; Turner 1973; Reynolds and Wilson 1991), the process entailed in transcribing an ancient text remains opaque. Papyrology is, in essence, a 'self consuming labor which leaves little or no trace of itself' (Youtie 1963: 11) and the expertise of papyrologists, as with the expertise of any professional, is a valuable but surprisingly elusive resource. There are only a few discussions by papyrologists themselves that analyse the nature of papyrology. Additionally, very little research has been done in the field of cognitive psychology regarding how experts may read damaged and deteriorated texts, or even texts in languages other than (clearly printed) English.

2.1.1. Papyrologists on Papyrology

There are a handful of papers written by those working on ancient documents which attempt to analyse and relay the nature of the processes used to generate satisfactory readings; two papers by the papyrologist H. C. Youtie, a guide to the reading of cuneiform texts by the assyriologist E. Reiner, a pamphlet regarding decipherment by P. Aalto, a Victorian guide to the reading of 'ancient' documents by E. Thoyts, a paper regarding the problems encountered in reading the Vindolanda ink and stylus texts by A. K. Bowman and R. S. O. Tomlin, a note in a review article by the papyrologist A. S. Hunt, and a discussion regarding papyrology by E. G. Turner in his texts on the reading and use of papyri.

Youtie's papers (1963, 1966) both attempt to describe the processes that the papyrologist goes through when transcribing a text—'What do papyrologists do?' (1963: 9). Youtie demonstrates that most accounts of papyri and the reading of papyri consider the studies to which they make a contribution and not the act of transcription itself, before focusing on the mechanics of papyrology—the publication, annotation, and editing of transcriptions of ancient texts. Youtie then goes on to attempt to describe what happens when he

is reading such an ancient text: 'A Theban receipt is not written to be deciphered letter by letter; it presupposes a reader who has sufficient information to read intuitively' (1963: 14). Youtie's attempts at describing this activity (although at times hyperbolic to the non-practitioner) demonstrate the complex, recursive nature of the task:

He tries to take account of the text as a communication, as a message, as a linguistic pattern of meaning. He forms a concept of the writer's intention and uses this to aid him in transcription. As his decipherment progresses, the amount of text that he has available increases, and as this increases he may be forced to revise his idea of the meaning or direction of the entire text, and as the meaning changes for him, he may revise his readings of portions of the text which he previously thought to be well read. And so he constantly oscillates between the written text and his mental picture of its meaning, altering his view of one or of both as his expanding knowledge of them seems to make necessary. Only when they at last cover each other is he able to feel that he has solved his problem. The tension between the script and its content is then relaxed: the two have become one. (Youtie 1966: 253)

He also elucidates the difficulties involved in the process of reading such damaged and deteriorated texts:

the memory of all the effort that it has cost, the doubts, the hesitations, the numerous false starts and new beginnings, the guesses sometimes confirmed, sometimes rejected by the script, the continual recourse to books for information of every sort—lexical, grammatical, palaeographic, historical, legal; the every threatening awareness of . . . visual and intellectual inadequacy; the interludes of exhaustion and depression . . . (Youtie 1963: 17)

Ultimately, it is demonstrated that it is difficult to explain in the final annotated version of the text how the papyrologist got from papyrus to transcription and translation. The complex process of transcribing a text is lost in its documentation.

Aalto's pamphlet aligns the reading of ancient texts with the work of cryptographers, demonstrating how techniques such as transposition and substitution are used in both cracking military codes and reading ancient texts.¹ He demonstrates how different techniques can

¹ Aalto discusses how experts in ancient languages from the British Museum and elsewhere were successfully employed in the First World War as code breakers for the British army; e.g. A. S. Hunt, the Oxford papyrologist who excavated the Oxyrhynchus Papyri with B. P. Grenfell, worked with the War Trade Intelligence

be used to understand texts where the script, language, or both are unknown, and suggests the importance of statistical information, the generation of probabilities, and the comparison of grammatical and etymological elements when trying to decipher texts. Like Youtie, he stresses the recursive nature of the task: 'Every interpretation thus implies a long series of trials (and errors too!) before it can possibly be regarded as verified' (Aalto 1945: 18).

Reiner interestingly adopts 'the fiction of programming a machine to read cuneiform' (1973: 16) to sketch a very precise and detailed model of the process of identifying and understanding the written Assyro-Babylonian language. She shows that 'the reading of a Cuneiform text is based on information which includes all three components: syllabary, grammar, and dictionary...and also that the three components are interrelated, or in Saussure's words "tout se tient" in the system that we call language' (p. 15). The reading process is stratified into four sections (p. 7),

1. The basic value look up table
2. Finding the ultimate value
3. Segmentation
4. Morphosyntactic analysis

and Reiner talks of the 'generalized cross reference' between these sections (p. 22) needed to reach a conclusion regarding the text. However, although these steps may be extrapolated to cover the reading of other languages, the remainder of the paper focuses exclusively on issues regarding particular aspects of cuneiform, such as 'the features of voice and emphasis in stops and sibilants' (p. 25). Whilst providing an interesting introduction into how to approach such a complex language system, this paper is mostly a comparative analysis of contemporary research on Akkadian phonetics, phonemics, and morphology.

The one book which purports to explain 'How to Decipher and Study Old Documents: Being a Guide to the Reading of Ancient Manuscripts' (Thoyts 1893) details how to gain access to and place into historical context 'ancient' texts, such as parish registers, deeds,

and monastic charters (it is worth noting that there were hardly any classical cursive documents available for transcribing in 1893). However, Thoyts shows little insight into how one may actually *read* such texts: ‘Written in a language I knew not, relating to customs no longer existing, all was strange and unfamiliar. I toiled on, by degrees light dawned and the difficulties melted away’ (p. xi). Thoyts’s conclusion indicates the paucity of advice she gives regarding how to approach and reason about documents: “‘Persevere and Practice’ is the best motto I can give to those interested in the matter, for proficiency comes quickly to those who seek it, and, as in all subjects, ‘Nothing succeeds like success’” (p. 143). Although an intriguing text regarding the growth of antiquarianism in the late nineteenth century, this guide offers little in the understanding of how experts read, transcribe, and understand ancient, often damaged, texts, never stretching beyond the obvious: ‘a transcriber’s work properly consists chiefly in correctly putting into modern handwriting the deeds which are only illegible to the uninitiated’ (p. 9).

In contrast, Bowman and Tomlin’s paper (2005) details the experiences they have had over the past quarter century or more in reading damaged and abraded texts, particularly the wooden stylus tablets from Vindolanda, providing the most comprehensive illustration of the exercise to date. They provide some introspection and analysis of their own working processes:

First we identify the shapes of letter forms, which fall into a range of types and different hands, then we read individual letters and combinations of letters to the point where we can construct words, phrases, sentences and finally whole texts. But of course we do not normally transcribe letter by letter in a completely neutral and automaton-like fashion, and then realise that we have transcribed a word or sentence; there is a point, very quickly or perhaps immediately reached, at which we bring into play our corpus of acquired linguistic, palaeographical and historical information which in effect predisposes or even forces us to predict how we will identify, restore or articulate letters or groups of letters. And this is a recursive process. We do the easy bits, then make hypotheses about the problematic bits and test them in the context which we think we have established by what we can read. (p. 8)

An account of how they read and formulated a meaning of Vindolanda Tablet 974 is presented, giving an example of this process. They

show how the conclusions they draw rely on parallels with other texts, and how the accumulation of textual knowledge from other similar documents can improve their reading of each (the expanding knowledge base regarding Vindolanda aiding the reading of each subsequent text encountered). Reading an ancient document is then 'not a letter-by-letter transcription like sleep walkers, but a wakeful testing of possibilities in the light of other knowledge' (p. 12).

The importance of this cumulative knowledge is confirmed by an aside in a review article by the papyrologist A. S. Hunt, who notes that 'Successful decipherment is largely a matter of practice, and a rapid perception of what can or cannot be right is the product of ample experience': as a papyrologist successfully reads more documents, his ability to read documents improves (1928: 24).

Another description of the nature of the process of papyrology is to be found in a text by Turner (1968), and whilst he, again, describes the mechanics of the papyrology in describing the Leiden system² and publishing formats, he provides a romantic³ view of the actual act of transcribing and translating such ancient texts:

What does a papyrologist do? He is engaged in eliciting new knowledge and doing his best to guarantee the reliability of that knowledge. His proper and professional field of competence is in relation to the text themselves and what they say. In other fields, such as Greek literature, Roman history, law etc., though his views may be listened to with respect, he will be an amateur. Curiosity and excitement will have started him on his quest; a passion for truth will bring him to its conclusion. (Turner 1968: 73)

It is this 'curiosity and excitement' that make the demanding process of reading a text worthwhile. In an early publication about the texts from Vindolanda, Bowman and Thomas comment on the rewards of such effort: 'Arduous and time consuming though the task may be, it is one which has great compensations. It is exhilarating to feel that one is penetrating the mind of an ancient man and gradually making sense of the ink marks he put down on wood two thousand years ago'

² A series of symbols which denote various characteristics of the original text, see 2.5.3.

³ For the act of reading and translating ancient documents as wistful subject matter for the arts, see William Wordsworth's poem 'Upon the Same Occasion' (1819), Tom Stoppard's play *The Invention of Love* (1997), and Lesley Saunders's poem 'The Uses of Greek' (1999).

(1974: 22). This echoes the passion Turner shows for papyrology in his books where he sets out ‘what is useful for a man to know before he begins to use ancient texts’ (1980: 1), such as writing materials, the rediscovery of papyrus, and what papyrologists attempt to do when editing a papyrus (showing much acknowledged debt to Youtie’s discussions); and gives examples of how he constructed readings of various texts (1973). He describes papyrology as

the process of discovery and exploitation of a new piece of information, won unexpectedly from written relics of the past. The pleasure, indeed thrill, or discovery is the reward for a great deal of drudgery: meticulous attention to exact setting out of what the papyrus contains, step-by-step testing of the hypothesis on which even simple restoration is undertaken. This I take to be the meaning of editing, and this is the proper task of a papyrologist. He will not do it well unless he attempts a further step—to reconcile what is new with what was already known. (1973: 14).

This passion can be neatly juxtaposed with real-world concerns, however: A. S Hunt notes, in an article for the *Oxford Magazine*, that whilst ‘the unending variety is one of the charms of the subject ... the fact must be faced that, as with archaeological work generally, the pecuniary reward is not likely to be large. Papyrology is no road to an early marriage, or even to the possession of a motor car.’ (1930: 86).

2.1.2. Psychology and Papyrology

How papyrologists read ancient texts, as well as being seldom addressed by the experts themselves, has not been the focus of any psychological research. The field of psycholinguistics is dominated by more mainstream questions regarding language (for example, the acquisition of language, the relationship of linguistic knowledge to language usage, and the comprehension and production of speech (Aitchison 1998)) rather than the mechanics of such a complex task. Although there is a growing interest in translation studies⁴ in the

⁴ The emergent field of translation studies (Venuti 2000) addresses questions regarding the production of ‘a target language version of the source language text’ (Danks and Griffin 1997: 164).

humanities, there has been little work done on the psychology of translation⁵ (papyrology differing from straight translation on many levels anyway, given that decipherment is a large part of the papyrology process.)⁶ The psycholinguistic studies which do exist regarding the use of a second language focus on the psychological, social, and educational issues surrounding bilingualism in individuals, or the learning of a second language. Studies on how expert readers access the separate lexicon of a second language is in its infancy (Hornby 1977; De Groot and Barry 1994). Most research done on linguistics concerns the acquisition and use of the English language, and the research that does exist 'suggests that differences may exist in the reading processes between languages with different writing systems, orthography, morphology, and syntax' (Asher 1994: 3461). This would mean that the large structural differences between, say, English and Latin,⁷ would render the reading of a Latin document by a native English speaker a very different process than the reading of

⁵ Cognitive psychologists have not explored straight translation thoroughly, let alone how experts deal with the reading of degraded and difficult ancient texts. This may be partly because they are not really aware of the emergent discipline of translation studies, or that those in the field who have recognized its relevance may have given priority to the study of simpler tasks regarding how humans read, of which there is scant understanding. Solutions to those may provide the answer to further questions regarding the cognitive processes involved in translation (De Groot 1997: 29).

⁶ All the surveyed research in translation studies assumes that there is no problem in the actual *reading* of the text whilst comparatively studying texts and their translations (Carr *et al.* 1994). Although some of the techniques used in translation studies can be adopted in analysing the procedures papyrologists use, the focus of the two disciplines is different, as papyrologists primarily aim to deliver a transcription of the text in its native language. As Youtie stated, 'we do of course translate the texts that we derive from papyri but . . . we must first obtain the texts' (1963: 9). Undoubtedly there will be some shared cognitive processes between the two fields, but translation studies has little to offer at present in understanding how papyrologists read ancient texts.

⁷ Although English derives many words from Latin, the structure of the languages differs widely (Ellegard 1963). English is much more dependent on word order than Latin, which utilizes word endings more as indication of case, Latin words reflecting their purpose by the altering of their structure (Morwood and Warman 1990: 46). The Latin concept of gender in language is entirely different from that of English with nouns being assigned gender-specific traits. Latin makes more use of noun and adjective stems and less use of adverbs than English. Latin uses an entirely different case structure than English (Conway 1923: 84–106). Put succinctly: 'You cannot apply the rules of Latin grammar to English' (Morwood and Warman 1990: 48).

an English text by the same reader. In turn, most current research on reading would therefore be inappropriate to this investigation.

Research which specifically addresses the resolution of ambiguity in texts, whether that is lexical (Hogaboam and Perfetti 1975; Swinney and Hakes 1976) or structural (Hirst 1987), concurs on one thing: that there is no agreement on the principles which readers use for disambiguation when reading texts. There has, however, been a great deal of research done on how readers carry out simple reading tasks, and this is covered in section 2.8.1.

2.1.3. Widening the Context

Ancient documents are not the only type of texts which require close study to provide primary source data for humanities scholars—documents from the medieval and modern periods can also demand the type of scrutiny discussed here to yield their information. Discussions regarding the history and reading of medieval texts can be found in Clanchy (1993) and Bischoff (1990), Parkes discusses the conventions used in the presentation, dissemination, and communication of medieval texts (1991), and punctuation (1992), and Munby *et al.* (2002) discuss conventions used in Tudor and Stuart handwritten documents. However, I have found no account that discusses the reasoning processes applied when reading a medieval or modern manuscript in the analytical way that papyrology has been discussed (which is not to say that it does not exist, however). There also exist many texts regarding the wider context of the history of writing systems and reading (Gaur 1987; Coulmas 1989; Healey 1990; Manguel 1997; to give but a few examples.)⁸

From these a wider context can be constructed. Writing, for the most part, consists of a set of symbols that the writer usually intends to be legible. These symbols must have elements which enable a reader to distinguish them from other symbols.⁹ Readers try to

⁸ An extensive bibliography regarding the history of literacy can be found in Clanchy (1993), Bischoff (1990), and Parkes (1991, 1992).

⁹ See Parkes (1991, 1992) for an introduction to the development of the graphic principles which determine letter and punctuation forms. Parkes describes these conventions as a ‘grammar of legibility’ throughout both texts, tracing the development

construct meaning from these symbols. The relation of symbol to signal, representation to understanding, is a complex one—and one not fully understood by those who carry out the task, or those who research it, as indicated above. The same problems can be echoed in the reading of a modern letter as in reading an ancient text.¹⁰ There can be a lot learnt from a close study of how particular experts read and construct meaning from particular texts, and use the knowledge they have gathered about the conventions of these texts to aid in their transcription: hence the focus of the research presented in the remainder of this chapter.

2.1.4. Papyrology Undiscovered

Although ancient manuscripts provide a primary source for historians, linguists, and others, until recently very little attention has been given to the process involved in deciphering such texts. When it has been discussed, no attempt has been made to make explicit the reading process, the accounts being discursive rather than quantitative. There has been no comparison made of the differences and similarities between the work of individual papyrologists. Ultimately, when the readings of ancient documents are published, there is a

of written conventions designed ‘both to improve the intelligibility of . . . scripts, and to facilitate access to the information transmitted in the written medium . . . This complex of graphic conventions became the process by which the written medium of communication functions. A written text presupposes an indeterminate audience disseminated over distance or time, or both. A scribe had no immediate respondent to interact with, therefore he had to observe a kind of decorum in his copy to ensure that the message of the text was easily understood. This decorum—the rules governing the relationships between this complex of graphic conventions and the message of a text conveyed in the written medium—may be described as “the grammar of legibility”’ (1991: 1–2). Further discussion of writing systems as symbols can be found in Clanchy (1993: ch. 8).

¹⁰ Bowman and Tomlin (2005) provide an account of an amusing example: ‘There is a story about another master of palaeography, M. R. James, who catalogued all the western manuscripts in Cambridge, but is now better known for his *Ghost Stories of an Antiquary*. His own handwriting, appropriately enough, was illegible. His colleagues once received an invitation to dinner: “. . . we *guessed* that the time was 8 and not 3, as it appeared to be, but all we could tell about the day was that it was not Wednesday.” As it happens, this was good palaeographical method.’ (Originally told by George Lyttelton in Hart-Davis, 1978: 14.)

focus on the texts themselves rather than the processes undertaken to read them:

The general accounts, the surveys, the reports, whatever they are called, tell us nothing about the work done by the papyrologist. They all take up where he leaves off. They talk about papyri as they are after the papyrologist has finished with them, when he has already completed his transcriptions, added his philological and sometimes historical commentaries, and made them available in learned journals or volumes of papyri to any specialists who have a use for them. (Youtie 1963: 11)

Youtie terms this documentation ‘public papyrology’. To be able to discover and describe the process of ‘private’ papyrology (‘what a papyrologist does privately, in the solitary confines of the library, in order to make public papyrology possible’, p. 11), detailed research regarding how individual papyrologists work was undertaken. ‘Knowledge elicitation’ techniques, frequently used in the fields of computing and engineering science to enable engineers to capture expertise and encapsulate it into computer systems, were used to elucidate this process.

2.2. KNOWLEDGE ELICITATION: A BRIEF GUIDE

The problem with trying to discover the process that papyrologists go through whilst reading an ancient text is that experts are notoriously bad at describing what they are expert at (McGraw and Harbison-Briggs 1989). Experts utilize and develop many skills which become automated and so they are increasingly unable to explain their behaviour, resulting in the troublesome ‘knowledge engineering paradox’: the more competent domain experts become, the less able they are to describe the knowledge they use to solve problems (Waterman 1986). Added to this problem is the fact that, although knowledge acquisition and knowledge elicitation¹¹ are becoming increasingly necessary for the development of computer systems,

¹¹ Knowledge acquisition is conventionally defined as the gathering of information from any source. Knowledge elicitation is the subtask of gathering knowledge from a domain expert (Shadbolt and Burton 1990).

there is no consensus within the field as to the best way to proceed in undertaking such research.

Discussions regarding how best to elicit knowledge for the basis of an expert system first appeared in the late 1970s (Feigenbaum 1977). Early attempts at eliciting, formalizing, and refining expert knowledge were so unfruitful that knowledge elicitation was labelled the 'bottleneck' to building knowledge-based systems (Feigenbaum 1977). Throughout the 1980s and early 1990s, protocols (often referred to as the 'traditional' or 'transfer' approach to knowledge elicitation) were developed regarding how a knowledge engineer should interact with a domain expert to organize and formalize extracted knowledge so that it is suitable for processing by a knowledge-based system (Diaper 1989; McGraw and Harbison-Briggs 1989; Boose and Gaines 1990; McGraw and Westphal 1990; Morik *et al.* 1993). These discussions centre on the suitability of different techniques (often derived from clinical psychology and qualitative research methods used in the social sciences) for the capture of knowledge, such as:

- Unstructured, semi-structured, and focused interviews with the expert(s).
- Think Aloud Protocols (TAPs), where an expert is set a task and asked to describe his or her actions and thought processes, stage by stage (see S. 2.3.2).
- Sorting, where the expert is asked to express the relationship between a pre-selected set of concepts in the domain.
- Laddering, where the expert is asked to explain the hierarchical nature of concepts within the domain.
- The construction of Repertory Grids, where concepts are defined by the way they are alike or different from other related concepts in the domain by comparing and contrasting values.

A discussion contrasting these with other knowledge acquisition techniques, highlighting their suitability regarding the elicitation of certain types of knowledge, can be found in Cordingley (1989).

Since the early 1990s the field has focused on automated knowledge elicitation, incorporating the same psychological techniques into computer programs, to make the interactions more productive, assisting, and in some cases replacing, the knowledge engineer

(White 2000). Such tools were, at first, implemented in a stand-alone, domain independent way, focusing on the collection of particular types of data. For example, the ETS and the AQUINAS systems (Boose 1990) are computerized representations of the Repertory Grid¹² method and have been used to derive ‘hundreds’ of small knowledge-based systems. Gradually, programs appeared offering implementations of various techniques bundled together as a knowledge elicitation ‘workbench’, such as the research prototype ProtoKEW (Reichgelt and Shadbolt 1992) which was later repackaged and marketed as the commercial PC-Pack system.¹³ Researchers have now started to utilize the internet, developing distributed knowledge acquisition tools such as RepGrid¹⁴ and WebGrid,¹⁵ which can be used remotely to build up knowledge bases and data sets. However, these computer tools produce the best results when applied to very small domains to build knowledge-based systems which carry out well-defined tasks, and are not successful at providing overviews of complex systems, or when used to describe domains about which little is known from the outset (Marcus 1988; White 2000).

¹² The repertory grid method is based on building up a matrix of information which defines a domain or subject. This is based on the theory of personality termed ‘Personal Construct Psychology’ (PCP) or ‘Personal Construct Theory’ (PCT) developed by the American psychologist George Kelly in the 1950s (Kelly 1955). Kelly derived a psychotherapy approach and technique called ‘The Repertory Grid Interview’ that aided his patients to discover their own ‘constructs’ regarding the relationship between items, people, or emotions, with minimal intervention or interpretation by the therapist. For example, border collies, dachshunds, and xoloitzcuintles are all breeds of dogs (and therefore individual ‘concepts’ or ‘elements’ of the domain called ‘dogs’). They have different characteristics such as height, weight, length of coat, intelligence, athleticism, rarity, cost, etc. These are ‘constructs’, in that they can be compared to each other on a numerical scale, or grid, to build up a profile of each type, or concept, of dog. In this way, an individual concept can be defined by comparing it to other concepts. The repertory grid was later adapted for various uses, including decision-making and interpretation, and is a common approach used in business and knowledge-based systems to represent the relationship between related items, objects, processes, or states (Shaw and Gaines 1983; Stewart 2004). An example of how this approach is used in relation to the Vindolanda texts is provided in section 3.2.3, where it is used to aid in the capture of data regarding letter forms.

¹³ <http://www.epistemics.co.uk/Notes/55-0-0.htm>

¹⁴ <http://www.csd.abdn.ac.uk/~swhite/repgrid/repgrid.html>

¹⁵ <http://tiger.cpsc.ucalgary.ca/WebGrid/WebGrid.html>

Although much research has therefore gone into the different individual techniques and tools which can be used to elicit knowledge from an expert, 'no single knowledge acquisition standard has emerged' as McGraw and Harbison-Briggs are keen to point out (1989: 52). There are actually few useful guides on how to undertake such an exercise from conception to completion, and how to choose the best selection of tools suited to the domain being examined (Dermott 1988). The knowledge engineer is left to try and ascertain how best to identify, collect, and rationalize any sources of knowledge, use any computational tools, interact with the experts, and develop and check any conclusions reached regarding the domain. External factors, such as amount of time available, amount of relevant external source material, and the availability of the experts, all affect the process. The scope of the research also affects the tools and techniques chosen: is the knowledge engineer trying to gain an understanding of an overall process, or an intricate understanding of a small task? Knowledge elicitation remains a complex, time-consuming process, and the choice of method and technique is specific to each domain investigated.

2.3. KNOWLEDGE ELICITATION AND VINDOLANDA

The primary questions to be asked in this research were: is there a general process that experts use when reading ancient texts? Can this procedure be elucidated? Additionally, what are the differences and similarities between individual experts' approaches to the problem? Also, although it has been noted that 'cognitive and knowledge-based problems . . . are in general common to ink texts and inscribed texts' (Bowman and Tomlin 2005: 8), how does the format and medium of a document affect the reading process? Finally, how does the documentation provided by the papyrologists on publication of a text relate to the actual reading of an ancient text, and can it indicate anything about the process?

A number of steps and observations were undertaken to gain an understanding of the general process the experts utilize when approaching an ancient text, and specifically, the Vindolanda ink

and stylus tablets. First, as with all knowledge acquisition tasks, the domain literature was researched, as reviewed above in section 2.1. Second, any other associated literature was collated. Although not a direct comment on the act of reading and transcribing, the two published volumes regarding the Vindolanda ink tablets contain detailed apparatus of the individual texts (Bowman and Thomas 1983, 1994).¹⁶ The standard publication format (the transcribed text, marked up using the Leiden¹⁷ system, followed by the critical apparatus: variant readings, misspellings, line by line comments and explanations, and the translation of the text), is all that is presented of the process which was undertaken in the reading and understanding of the documents.

As an example of this format, here is the published commentary of stylus tablet 836, by Bowman and Tomlin (2005: 10). In the text presented here, letters printed in **boldface** are those which can be read with confidence; letters printed in ordinary type are read with some measure of conjecture; underlinings indicate traces of letters which cannot be identified with confidence:

banus bello suo salutem
[traces only]
acc__erunt in in uecturas
de_arios octo reliquos solues
5 rios nouem qua__r_r__
sam dari debeb__
[interlinear addition?]
em libris
dus uale

Albanus to his Bellus greetings... they have received for transport costs 8 denarii. You will pay the remaining 9 denarii... ought to be given (?)... nine pounds (?)... Farewell.

Line 1: there is a trace between the first and second l in **bello** which might or might not be a letter. The scratches on the wood show that this overlies an earlier text.

¹⁶ A further volume, Bowman and Thomas (2003), was subsequently published, but this research was undertaken before that text was available.

¹⁷ See section 2.5.5.3.

Line 3: The correct reading is almost certainly **acceperunt**.

Line 4: The word at the end of the line presents particular difficulty.

Of the first three letters of **solues** only the **o** is certain. There is a clear high horizontal which has to be ignored if the first letter is read as **s**. The third letter might be **p**, and there is another apparent high horizontal which is discounted. The attraction of reading the word **solues** (from the verb **soluere** 'to pay') is obvious if the word '**denarios**' occurs twice in lines 4–5.

The apparatus aims to cover comprehensively the difficulties, reasoning, and alternative hypotheses regarding the final transcription. As such, it is the best representation of the different types of knowledge used in the reading of texts available without carrying out further investigation (although the sequential order of the different stages in their reading are lost due to the reporting format). These texts were obtained in digital format to enable in-depth research.

Three experts were then identified who were working on the ink and stylus texts, and who were willing to take part in this investigation. (They shall be referred to as Expert A, Expert B, and Expert C, so as to spare their blushes, and not to imply any criticism of their work: this investigation is concerned with asking how an expert operates, rather than making judgements on those operations.) Expert A works equally on the ink and stylus texts from Vindolanda. Expert B works mostly on the stylus tablets, and other similar incised texts from the period. Expert C's studies mainly focus on the ink tablets, and other similar texts from the period. All three experts are English male academics who have been working on such texts for over twenty-five years. They graciously gave their time and permission for this research. (It must be stressed that access to experts who are willing to contribute the time to allow their expertise to be studied is a prerequisite for the success of a project such as this.)

A series of investigations were carried out, utilizing knowledge elicitation techniques, particularly Think Aloud Protocols. The data captured provides explicit and quantitative representation of the way the papyrologists approach damaged and abraded texts. General procedural information was also collated.

2.3.1. First Stages in Knowledge Elicitation

The first stages of knowledge elicitation were fairly informal. The experts were observed whilst going about their tasks, and unstructured interviews were undertaken, where the experts described their domain, and the individual processes and techniques that they preferred. More structured interviews were then undertaken, when the experts were asked to describe particular facets of their task, such as the identification of letter forms (which is the focus of Chapter 3 in this book), and the role of grammar, word lists, and external historical and archaeological resources in the reading of the documents. Joint sessions, where Expert A and Expert B discussed their readings of the stylus texts, were also attended. These sessions were documented, and a broad understanding of papyrology was formed, whilst building up a good working relationship with the experts.

2.3.2. Think Aloud Protocols

However, a more formal approach was needed to build up some quantitative data on the reading of such texts. It was decided to undertake a series of Think Aloud Protocols (TAPs), a technique adopted from experimental psychology, where the expert is urged to utter every thought that comes to mind whilst undertaking a specified task. These sessions are recorded, transcribed, and analysed to allow the observer to identify the different steps characterizing conceptual processes more precisely. Although ‘an expensive and meticulous research method that has had its share of growing pains’ (Smagorinsky 1989: 475), the collection of verbal data in this manner has been a procedure used in the social sciences for three-quarters of a century (Duncker 1926). Protocol analysis has been shown to be ‘a very useful addition to the repertoire of research tools... The data from most other tools yield little about the internal structures of cognitive processes, particularly when the tasks are complex. Think-aloud protocols, in contrast, can yield significant information about the structure of processes’ (Smagorinsky 1989: 465).

Ericsson and Simon (1993) show that verbalization does not interfere with the cognitive processes discussed, and that there is

little difference between results whether the information is collected concurrently (whilst the expert undertakes a task) or retrospectively (when the expert details what they did whilst undertaking a task). Most cognitive studies of translation and interpreting use TAPs as the tool of choice (Danks *et al.* 1997: p. xv); for example, Kiraly (1997) effectively used TAPs to investigate how translators work.

The three experts were set various tasks to gain an insight into how they approach and reason about the ink and stylus texts. It was explained to each the nature of the Think Aloud Protocols, and how and why the experiment was going to be carried out. For the most part, these TAPs took place in the workplace of the experts, where they would usually work on such texts. Various data were collected:



Figure 2.1. Ink tablet III 616. Although photographed together as it was believed they were part of the same text on first inspection, these two fragments are actually from different documents, a fact which only became apparent during the process of reading the texts: the upper section is ink tablet III 684 (which appears to be an account of timber needed for building a granary: Bowman and Thomas 2003: 135). The photograph with both sections was given to the experts to read. Bowman and Thomas (2003: 79) published a transcription of the main body of the document: ‘all of you take good care to see that if any friends come they are suitably entertained’.

- All three experts were given a pack containing various images of tablet III 616 (see Figure 2.1), and asked to come up with the best reading they could at their own leisure. In the session they were asked to talk through how they arrived at their reading, and to describe the processes they undertook whilst coming up with their final text.
- All three experts were presented, on the day, with an image of an ink text (III 663). They had not been asked to prepare this text, and so discussed how they would go about tackling a text from the outset. (This text was identified as being one of the ink texts the experts would be least familiar with. In actuality, Expert A and Expert C had some prior knowledge of this text. Expert B had never seen this text before.) This provided data to compare how experts reason about texts they have had the time to prepare, and those they have had little experience with.
- Expert B was asked to talk through his reading of III 663 again at a later date, to describe how he had read the document. He had not looked again at the ink text in the interim. This was done to investigate how retrospectively talking about the reading of texts differs from the concurrent explanation of the process.



Figure 2.2. Ink tablet III 663. Clear handwriting in some parts of the text would allow the experts to try to produce an elementary reading of the text on first sight. This text is the first part of a diptych, which was subsequently published in Bowman and Thomas (2003:122): ‘has made (?), with which you agreeably (?) comfort me just as a mother would do. For my mind . . . this sympathy (?) . . .’

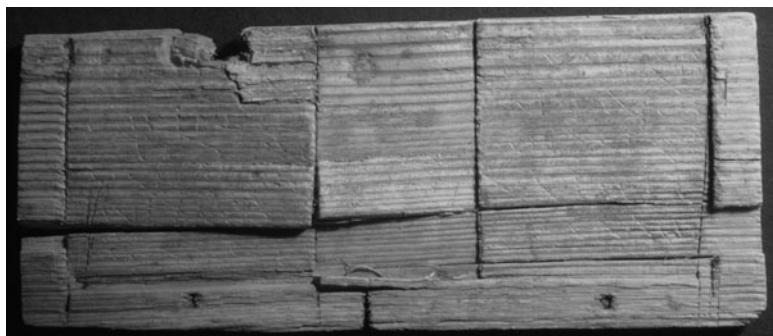


Figure 2.3. Stylus tablet 797, containing a single text written in two columns on either side of a central recessed strip. This was intended for the wax seals of the witnesses to a legal document, but the tablet was evidently not used for this purpose, or has been reused. A partial reading of the last two lines of column 1 on the left and the five lines of column 2 on the right is published in Bowman and Tomlin (2005): ‘The text itself is a letter, or rather the end of a letter, in which the writer offers a guarded apology to the recipient: “... angry with me. [...] sometimes I am surprised at why you are now angry with me. Farewell.”’

- The experts who deal with stylus texts (Expert A and B) were given various images of two stylus texts, 974 and 797, from which they tried to obtain readings. They were asked to discuss these in full, to provide data on how reading the stylus tablets differs from reading the ink tablets.

These discussions and TAPs sessions were recorded, and transcribed verbatim, giving a large set of data (23,000 words of discussion; 16,000 words pertaining to the reading of the ink tablets, 7,000 regarding the reading of the stylus tablets).

2.4. ASSOCIATED TECHNIQUES: CONTENT AND TEXTUAL ANALYSIS

The majority of the data collected during this investigation was textual, due to the standard way of documenting Think Aloud Protocols, and the format of the Vindolanda text apparatus. Various

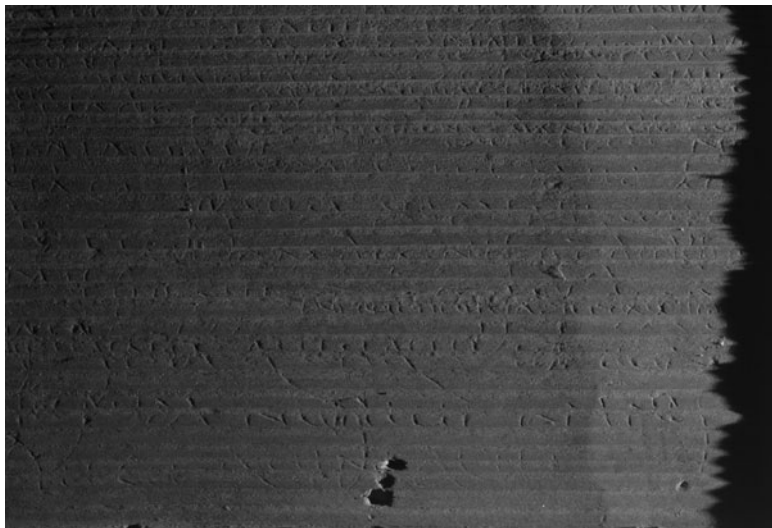


Figure 2.4. Stylus tablet 974. The inscribed text is confused by discoloration, woodgrain, and traces of at least one other text, which may or may not have been aligned with it. A partial reading was published in Bowman and Tomlin (2005). The text seems to be a letter, or official document, which (so far!) has been read as: ‘... of the Batavians (?) ... my fellow-citizen of the Bellovaci [name and verb lost] a slave called Verecundus, citizen (?) at Amiens. And I have given permission and travel-expenses (?) ... thirty-five; and I have kept that slave fifteen years.’

techniques have been developed in the social sciences and the humanities to analyse such data, the two most important being content and textual analysis.

2.4.1. Content Analysis and Vindolanda

Content analysis, a ‘method of studying and analyzing written (or oral) communications in a systematic, objective, and quantitative manner’ (Aiken 1971: 433) is an unobtrusive, context-sensitive, empirical process in which texts are reduced and condensed into a numerical format in order to estimate some phenomenon in the context of data (Holsti 1969; Krippendorff 1980; Stemler 2001). It has been used extensively since the 1930s in the analysis of large volumes of data, such as text-

books, comic strips, speeches, advertising, and the psychological analysis of personal communication, and is a large field in communication studies. Although content analysis has its problems with reliability and statistical validity (Franzosi 1990), has been accused of being subjective (Anderson 1974), and can be time-consuming and laborious (Aiken 1971: 279), it remains the most thorough and useful way to gain an empirical insight into the structure and content of complex textual sources, and is the standard technique used in encoding and analysing the data from Think Aloud Protocols (Ericsson and Simon 1993). Texts are divided into defined units¹⁸ and labelled, and these units can be used as a basis for statistical analysis.

In the case of the data from Vindolanda, the *subject* of the linguistic unit was identified as the defining feature, and the text split into sections where the subject of the phrase, sentence, or sometimes paragraph, changed. A number of pilot studies were undertaken before settling on an innovative encoding scheme which comfortably encompassed the data from the apparatus, transcripts, and preliminary knowledge elicitation exercises, resolving the different types of knowledge presented into an overall framework. The final, novel encoding scheme is presented in Table 2.1. It resolves the reading process into a finite set of ‘modes of thought’, or ‘states’, which can be used to track the development of the reasoning process.

Table 2.1. The encoding scheme resolved from an analysis of the Vindolanda textual data.

Reading level	Thematic subject
8	Meaning or sense of document as a whole
7	Meaning or sense of a group or phrase or words
6	Meaning or sense of a word
5	Discussion of grammar
4	Identification of possible word or morphemic unit
3	Identification of sequence of characters
2	Identification of possible character
1	Discussion of features of character
0	Discussion of physical attributes of the document
–1	Archaeological or historical context ^a

^a Presented as ‘–1’ to mark the fact that experts are explicitly referring to other sources and not only this document.

¹⁸ There are many different ways in which to break down a text in content analysis, such as word, sentence, paragraph, item, or theme (Holsti 1969: 180).

The Think Aloud Protocols were encoded in this manner, the time spent (in seconds) discussing each different subject also being recorded as well as the amount of words spoken.

A selection of critical apparatus from *Tabulae Vindolandenses II* (Bowman and Thomas 1994) was also analysed. It was important that the critical apparatus which were analysed contained enough information (some of the published texts are very fragmentary) to be useful, and contained the type of text that would be relevant; there was no point in comparing official army documentation with, say, literary texts. For this reason the texts were not selected randomly¹⁹ from the publication, but the authors were consulted to see which texts would be most significant in relation to the task. Knowledge elicitation techniques have shown that asking experts to recall the five or ten most important instances of their work can yield the best examples of the process (Morik *et al.* 1993), and so one of the experts was asked to point out the most relevant and important texts from their publication. Seven of the critical discussions were analysed. It is acknowledged that the purpose of the critical apparatus is to represent the meaning of the text, and difficulties associated with meaning, rather than to show exactly how the document was read, but it was hoped that analysis of these texts would provide some interesting insight into the type of information discussed by the papyrologists.

Because of the nature of the research there was only one person encoding the texts, leading to problems with validity. To increase the reliability of the analysis each text was encoded twice, and the two resulting spreadsheets compared to highlight any problematic areas, which were then reanalysed to make sure that the system was consistent and therefore reliable. All texts were encoded before any analysis of the data took place. Such techniques are the best way to ensure the validity of results when relying on one researcher for the content analysis of texts (Stemler 2001).

2.4.2. Textual Analysis and Vindolanda

The textual data from both the Think Aloud Protocols and the published commentaries were also analysed using a common corpus

¹⁹ It can be argued that the selection of texts that are published are quite statistically random due to the nature of their survival.

linguistics program for lexical analysis, WordSmith,²⁰ in order to look for patterns in the language used whilst discussing the documents. Such tools have been used in corpus linguistics since the 1970s (Sinclair 1991; Lawler and Dry 1998; Hockey 2000), and can be used for both quantitative and qualitative linguistic investigations (Popping 2000). The generation of word lists, frequencies, collocates, concordances, and key words can 'unpack the political, social, and cultural implications of texts' (Hoey 2001: 3). For example, textual analysis has been used to investigate the occurrence of sexist language in children's literature, and how language spoken in the court room can affect the outcome of a case (Stubbs 1996). In respect to Vindolanda, these tools were used to look for linguistic differences between the experts, and to analyse the order in which they undertook discussions. The differences in language usage regarding the different tasks set was investigated, to see if this could aid in the investigation of how the experts work. These tools were also used to highlight information regarding different individual letter forms, as discussed in section 3.2.1.

No automated knowledge elicitation tools were used in trying to gain an understanding of how the papyrologists worked. This was because the task investigated was too amorphous to be suited to any of the tools currently available, which require well defined parameters regarding small domains to be successful (White 2000). However, a Repertory Grid tool was used to collect and analyse data regarding the letter forms of Vindolanda, as discussed in section 3.2.3.

2.5. RESULTS

2.5.1. The Technique of Individual Experts Compared

It is perhaps unsurprising that, superficially, the three experts used in this research seem to read documents differently (see Table 2.2). The individual tools and techniques they prefer differ greatly; Expert A makes most use of digital images and PhotoShop to examine the

²⁰ <http://www1.oup.co.uk/elt/catalogue/Multimedia/WordSmithTools3.0/>

Table 2.2. Comparison of different individuals from transcripts of discussions.

Expert	Average word count per document	Average time per document (seconds)	Average words per second
A	2187	417	5.24
B	3205	1109	2.89
C	1622	947	1.71

texts, Expert B favours drawing his own representation of the text as he reads the artefact, Expert C relies mostly on photographs. The three experts spent various amounts of time discussing the texts, and the word counts of these discussions also varied greatly. The speed which they talked differed a great deal, with Expert B and C more prone to periods of silence whilst they considered the documents. Expert B imparted almost double the amount of information regarding a text than Expert C; Expert A talked quickly (and energetically) whilst Expert C chose his words with precision. These results suggest that the experts work differently but it is subsequently shown that there are underlying similarities to their techniques.

Each individual also had his own linguistic style in discussing the texts, which can be seen in the differences in language that they use, and illuminated by providing examples of their reading of the word *adfectum* in tablet III 663.

Expert A described readings that he was very certain of, and described the final conclusion of the texts firmly. The words which he used most often which differentiated him from the other experts were CLEARLY, PROBABLY, CLEAR, LOOKING, ABLE, HAVING, REFERENCE, FOLLOWING, and REMEMBER, showing a very concrete approach to demonstrating what he could actually see in the document. For example:

There is a reference to AFFECTUM. And then ANIMUS MEUS, and clearly it's talking about emotional matters. The last two lines read HUNC ENIM AFFECTUM ANIMUS and then something. (Expert A, III 663)

Expert B was comfortable talking about the reasoning process he went through, and could detail very closely the problems he encountered. He was much more likely to raise different possibilities and conclusions from the smallest change in letter forms, quickly changing hypothesis (akin to the 'rapid prototyping' model of soft-

ware development²¹). The words he used most often were SORT, SEEMS, SUPPOSE, SEQUENCE, HYPOTHESIS, and ASSUMING, indicating that he was concerned with the generation of different conclusions as he read a text:

A letter to be sure either a B or a D followed by a FEC. So it is pretty certain A D F E C, ADFEC. Which is a verbal part of ADFECERE, to affect. Then one would either see if it continues into the next space, and the next line is pretty clearly TUM. So we have ADFECTUM (Expert B, III 663).

Expert C was much more cautious in making assumptions, and details his hypothesis by stressing the uncertainty of his readings at all times. The words he most often uses which differentiates him from the others are IMMEDIATELY, SUGGESTS, WILL, VERB, INSTRUCTION. Again, this indicates a fairly certain attitude to the discussion of what he can and cannot see, and also the fact that he discusses grammar much more than the other two experts:

So then we go straight away and read the third line. You can straight away read ADFECTUM. And that's alright, masculine singular accusative so that's easy enough. What ADFECTUS means . . . it's a pretty vague word, and it refers to various sorts of moods and possibly it may refer to physical illness or mental illness or something like that. Depression. Anyway. That's not a matter of reading, the problem is a matter of reasoning, ADFECTUM is a problem word. The reading is clear, HUNC ENIM ADFECTUM. (Expert C, III 663)

The order in which experts discussed key features of the documents and identified constituent words also varied greatly. For example, whilst discussing tablet III 663, the identification of the words in the document happened in very different orders. This can be illustrated by plotting where three words from the text, *adflectum*, *fecit*, and *qua*, were used (and therefore identified) in the course of the discussion (see Figure 2.5).

It can clearly be seen that the experts identify the words in different orders (Expert A: *adflectum* → *fecit* → *qua*, Expert B: *fecit* → *adflectum* → *qua*, Expert C: *qua* → *adflectum* → *fecit*) and at different times in the discussion.

However, this difference in order is hardly surprising when it becomes clear that when an individual expert is asked to discuss a

²¹ Where a working model of a software system is rapidly implemented to demonstrate the feasibility of the function, with the prototype being later refined, adapted, and improved for inclusion in a final product (Connell and Shafer 1989).

text on two separate occasions, the order in which he discusses the content varies considerably (although there does not seem to be any order to this: he does not present them in a linear fashion as they appear on the document on the second reading). Reading one is *fecit* → *qua* → *adfectum*, reading two is *adfectum* → *qua* → *fecit*, when they appear on the document as *adfectum* → *fecit* → *qua* (Figure 2.6).

Although on the surface, the three individual papyrologists seem to discuss these texts in different manners, using different techniques, and in different orders, the readings which emerged from these different sessions were, for the most part, converging towards the same conclusions, and highlighting the same areas of difficulty. For example, compare the three experts’ readings of III 663:

And we have got something dot MAE...FECIT QUA ME something
CUNDE CONSOLARIS SIC UT MATER ET HUNC ENIM ADFECTUM.
(Expert A)

FECIT QUA ME...and then in the next line what I was reading as
FACUNDE but actually may be ECUNDE...And then this CONSOLA
word, And then another letter, And then what looks like UT

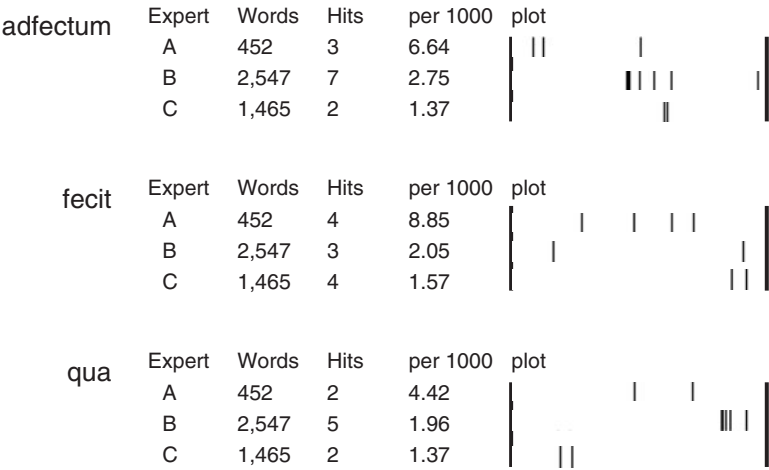


Figure 2.5. Occurrence of key words of document in discussion of ink tablet III 663 by three experts. ‘Words’ represents the number of words in the discussion, ‘Hits’ the number of times the term was mentioned. The plot function indicates where in the entire discussion the word occurred.

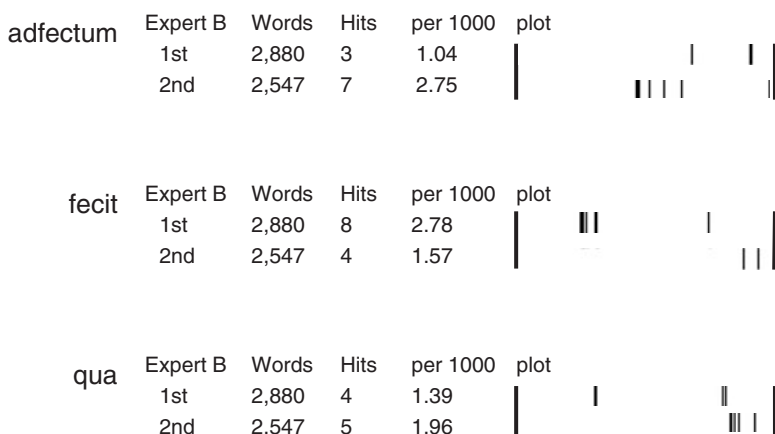


Figure 2.6. Occurrence of key words of document in discussion of ink tablet III 663 by the same expert on two separate occasions.

MATER... Then what I think is this sequence NUNC ENIM ADFLECTUM ANIMUS ME. (Expert B)

FECIT... QUA ME... before that we have got MA possibly MINA... So you then pick up the individual letters C U N D E... that could be CONSOLARIS... SICUT... And then MATER looks good, Then this could be something from the verb FACEO... This... reading is clear, HUNC ENIM ADFLECTUM... ANIMUS. And then you have M E and what looks like part of a U. (Expert C)

All three experts show the same confusion surrounding the reading of the word(s) prior to FECIT, the word which may be CUNDE, and the letters around SICUT. All clearly read FECIT, had some difficulties with the sequence QUA ME, and wondered about the possible meaning of ADFLECTUM whilst clearly reading the words around it. Further analysis of the Think Aloud Protocols, utilizing content analysis techniques, revealed hidden similarities between the experts' processes which illustrate the similarities in the way they reason about the texts.

2.5.2. The Cyclic Reasoning Process

A theme develops throughout the literature regarding papyrology: Youtie's 'oscillation' between different stages, Aalto's series of trials and errors, Reiner's 'interrelated' components and 'generalized cross

referencing', Bowman and Tomlin's 'recursive process', Hunt's 'rapid perception of what can and cannot be right', Turner's 'reconciliation' and 'step-by-step testing of the hypotheses', and even Thoyts's 'toiling on by degrees', indicating a cyclical process in transcribing and understanding the text under scrutiny. This can be shown where the experts discuss the overall process they use when transcribing a text. It is not a process of transcribing letter by letter (as Youtie, Aalto, and Bowman and Tomlin noted), but rather of proposing hypotheses and reasoning about these as more information comes to light:

Some people when faced with something like this will start by saying, 'Well we can identify 1,2,3,4, lines of writing here,' and start at the beginning of the first line and work their way through til the end of the last line and I suppose the idea is that you get some sort of objective view of the writing or build up of the text, letter by letter, but I don't work that way, I never have, and I don't believe that it really works very well. So what I have actually done, as I do with all these things, is really go to the point at which I think I can identify some of the letters in some of the words to start with and that begins to give me some sort of clue. (Expert A, III 616)

I suppose what I would do would be to try and do a letter by letter transcription and start seeing if anything was making sense... this is going to be a series of interlocking hypotheses which don't necessarily resolve themselves. In a way one is doing what one often does do, which is to say at this point you are sort of on the hypothesis, but you can't really be sure that it is so until you find something that kind of makes such obvious sense, that it must be right, whereas the first two lines or so do seem to work and are self contained. I don't know if you ever have tried life drawing, but it's often that you draw part of the figure and it all fits beautifully. Then you find maddeningly that the foot or something is too close, and you are kind of doing what requires a strength of character which is to redraw the good bit, to make it fit in with the other bit, otherwise you lose the proportion, the relationship of the whole, so that is always the problem in reading a text; how long you hold onto something that you are certain of, if it just won't fit in with anything else. (Expert B, III 663)

I could make out individual letters at first. What we're trying to do, or what I'm trying to do, is get words to make sense... not individual letters... (Expert C, III 616)

The identification of the core subjects covered in these discussions (Table 2.1) shows a novel scheme that can be used to illustrate the fact that discussions regarding texts vacillate between the identifica-

tion of features, letters, and words, and the production of meaning regarding these components. When plotted over time it becomes obvious that the reasoning process is far from linear, and depends on a complex cycle of interlocking elements. This can be seen in Expert A's discussion of ink tablet III 663. Similar graphs result when any of the discussions are plotted in this manner (see Figure 2.7).

Expert C begins by drawing some conclusions about the meaning of the document (level 8) before looking its the physical attributes (level 0). He then discusses what could be possible features of the text (level 1), before noting more physical attributes of the document (level 0). He then produces a word (level 4), looks at the characters within this word (level 2), and revises his initial word. Checking of the features (level 1) leads to identification of a character (level 2), the noting of a possible word (level 4) and a discussion of meaning of that word (level 6). In this manner the expert vacillates between the different levels in reading a document, until a resolution is reached regarding the sense of the document (level 8), or until he has exhausted all possibilities regarding the text. An extract from this

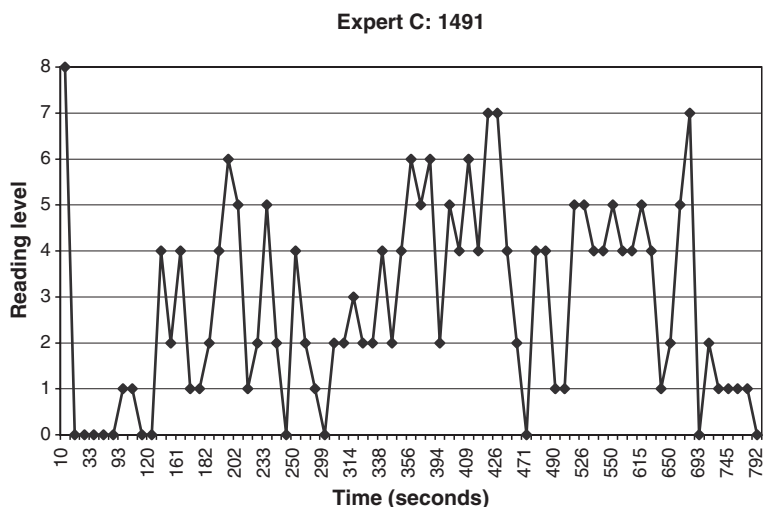


Figure 2.7. Discussion of tablet III 663 by Expert C, plotted by subject matter over period of discussion. It is obvious that the discussion oscillates between lower and higher levels of reading until a meaning is reached, and the problem is solved.

discussion, illustrating how the reading levels are appropriated, is presented in Table 2.3.

All the experts' discussions regarding the texts followed a similar pattern, with various hypotheses concerning the identification of features, characters, and words, being checked against other information from the document, until some resolution of ambiguity was reached. Modelling this process computationally, by implementing each different level as an agent, and using MDL as a means to pass information between levels (see S. 1.6.3, Chapter 4, and Appendix A), should provide a way to construct a tool to aid the papyrologists in their reading, as it would enable this process to be replicated.

2.5.2.1. *The order of the reasoning process*

It is assumed that readers inhabit particular modes of thought, or 'states', as they read a document, with one state transitioning to

Table 2.3. Excerpt from Expert B's discussion of III 663 illustrating the appropriation of reading levels to the text, and demonstrating the flow of the discussion between different levels.

Timing (secs)	Transcribed text	Reading level
131	Then, when we start reading it, well this looks like a very good Latin word FECIT—F E C I T.	4 Identification of word
139	And then immediately at the end what looks like Q U A N E.	2 Identification of characters
161	Which is a bit of a problem . . . if that is the end of the line is has to be QUA ME instead of QUAN, You wouldn't really expect to see it there.	4 Identification of word
167	But immediately when one gets as far as this you can see that this stoke coming down is E.	1 Identification of feature
174	So it's a descender from the line above, so we haven't got the top of N.	1 Identification of feature
182	Before that we have got M A possibly M I N A.	2 Identification of characters
186	Which obviously works with the Latin word DOMINA.	4 Identification of word
190	A woman has done something.	6 Discussion of word meaning
202	QUA is then better, the woman is then in the perfect.	5 Discussion of grammar

another as ambiguities are resolved: for example, deciding on a particular sequence of characters, which leads the reader to identify a possible word. It is possible to work out the latencies between different stages in the process (data from Think Aloud Protocols often being used to predict the sequence of cognitive processes executed in a given task: Ericsson and Simon 1993). This illustrates the likelihood of particular levels following on from each other, showing the order in which these processes occur. An attempt was made to see if there were any patterns occurring, resulting in some complex data. This is best represented in a general format as shown in Table 2.4, where the likelihood of one topic following another in the discussions is plotted on a grid. A 'high' probability means it occurs in over 5 per cent of the total discussions; likely, between 3 and 5 per cent of the discussion; often, between 1 and 3 per cent, low, between 0.05 and 1 per cent; unlikely, beneath 0.05 per cent of the total discussions.²²

The interesting factor in this table is that only the Word level, level 4, seems to have common connections with all of the different levels (save the archaeological and historical context level, which didn't occur during any of the TAPs). The other commonly mentioned subjects, features (level 1) and characters (level 2) (see Figure 2.4) seem to relate closely. For example, discussions concerning features of characters are most often directly followed by discussions regarding other features and characters. The word level is the only one where the next probable discussion is spread fairly equally across the range of topics available.

This is interesting when compared to an initial 'subject-object' analysis of the critical apparatus from *Tabulae Vindolandenses II*. Although these apparatus do not relate the order in which individual sections of the texts were read, it is possible to categorize the data that is presented both into the encoding scheme described, then into a more complex inner scheme which looks for any connections between the subjects by noting any secondary reference to other subjects. For example, the expert may be noting a difficulty about the identification of a word, which depends on a problem with the

²² It is worth noting that just because discussions directly follow on from each other they may not be directly related in some cases. However, for the most part, as the experts were talking exclusively about their reading of the texts, the order of the discussions should indicate something about the order of their thought processes.

Table 2.4. Likelihood of one reading level following another sequentially in the discussions.

1st topic of discussion	2nd topic of discussion									
	–1. Archaeol. info	0. Physical	1. Feature	2. Character	3. Char. sets	4. Word	5. Grammar	6. Word meaning	7. Phrase meaning	8. Meaning of doc.
–1. Archaeol. info,	Unlikely	Unlikely	Unlikely	Unlikely	Unlikely	Unlikely	Unlikely	Unlikely	Unlikely	Unlikely
0. Physical	Unlikely	Often	High (M)	High	Low	Low	Low	Low	Low	Low
1. Feature	Unlikely	Often	High	High (M)	Often	Likely	Low	Low	Low	Low
2. Character	Unlikely	Often	High	High (M)	Often	High	Low	Low	Low	Low
3. Char. sets	Unlikely	Low	Low	Often	Low	Often (M)	Low	Low	Low	Low
4. Word	Unlikely	Often	Likely	Often	Low	High (M)	Often	Often	Likely	Often
5. Grammar	Unlikely	Low	Often (M)	Low	Low	Low	Low	Low	Low	Low
6. Word meaning	Unlikely	Low	Low	Low	Low	Low	Low (M)	Low	Low	Low
7. Phrase meaning	Unlikely	Often (M)	Low	Often	Unlikely	Low	Low	Low	Low	Low
8. Meaning	Unlikely	Often (M)	Often	Often	Low	Low	Low	Unlikely	Low	Low

(M)=most frequent occurrence on that level. This is a ‘state transition probability matrix’, showing how the reader changes from one mode of thought into another as they read a text.

characters within the word, giving a main subject of word, and secondary subject of character. Similar subject/object models, or 'semantic triplet' (subject-action-object) models, have been recently developed and adopted by others in the field of content analysis, such as Franzosi's analysis of narrative structure (Franzosi 1997). By using such a method it is able to illustrate the transient relationships between different types of knowledge, seeing what type of knowledge the different instances depend on to reach their conclusions. The patterns which emerge from within the critical apparatus of the Vindolanda Texts also indicate that the word level is the most linked of all the topics discussed (Table 2.5).

While the lower level processes, such as the identification of features and characters, mostly relate to each other, and the upper levels, such as discussion of word meaning and meaning of document, mostly relate to each other, it is only the word level which shows a relationship with all the different types of information discussed.

The identification of words, then, seems to be the topic which is most likely to relate to all the different types of information used within the process of reading an ancient text. Although not the most commonly discussed topic (see Figure 2.4) it appears to be the most flexible, deriving identification of word units from various sources. Perhaps this suggests that it is the most basic and central of the whole process, but this needs to be the focus of further research: the data sets here are rather small. The relationship between different reading levels in the process of reading an ancient text may serve as a useful starting point for further statistical analysis in the future. A general summary of the discussions most likely to precede and follow each level can be found in Table 2.7.

2.5.3. Something to Talk About

By categorizing the discussions in the manner discussed above, it is easy to illustrate which different topics the experts talk about most when describing their reading of a text. There seems to be a general trend in all the discussions (Figure 2.8).

All three experts mostly talk about the features of the characters, the identification of characters, and the identification of individual words throughout the discussions. In comparison, the discussion of

Table 2.5. Relationship of main and subsidiary topics within the Vindolanda apparatus.

Main topic	Secondary information									
	–1. Archaeol. info	0. Physical	1. Feature	2. Character	3. Char. sets	4. Word	5. Grammar	6. Word meaning	7. Phrase meaning	8. Meaning of doc.
–1. Archaeol. info,	High (M)	Unlikely	Unlikely	Unlikely	Unlikely	Unlikely	Unlikely	Unlikely	Unlikely	Unlikely
0. Physical	Low	High (M)	Likely	Low	Unlikely	Unlikely	Low	Unlikely	Unlikely	Unlikely
1. Feature	Unlikely	Likely	High (M)	Likely	Low	Unlikely	Unlikely	Low	Unlikely	Unlikely
2. Character	Unlikely	Likely	Likely	High (M)	Likely	Likely	Unlikely	Low	Low	Low
3. Char. sets	Unlikely	Unlikely	Low	Low	High (M)	Likely	Likely	Likely	Unlikely	Low
4. Word	Low	Low	Low	Likely	Likely	High	High	High (M)	Likely	Low
5. Grammar	Low	Unlikely	Low	Unlikely	Low	Low	High (M)	High	Low	Low
6. Word meaning	Likely	Low	Unlikely	Unlikely	Low	Low	High	High (M)	Low	Low
7. Phrase meaning	Likely	Unlikely	Unlikely	Low	Likely	Low	Low	Likely	High (M)	Likely
8. Meaning	High	Unlikely	Unlikely	Low	Low	Low	Low	Low	Low	High (M)

(M) most frequent occurrence on that level.

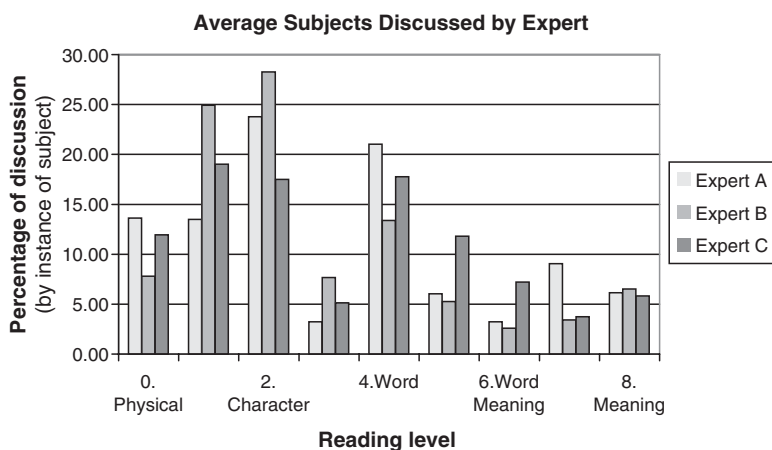


Figure 2.8. The average amount of subjects discussed by each expert throughout the discussions, by percentage of the discussion by instance.

the meaning of the document as a whole actually takes up a small part of the process, as do discussions regarding word and phrase meaning, grammar, and the identification of character sets. Although there are some differences between the distribution of the data between experts, this figure shows the predominance of discussions regarding features, characters, and words in the reading of ancient texts above that of grammar, phrase meaning, etc.

Conversely, the discussions regarding the identification of characters and words are most often the shortest amongst those occurring in the transcripts, being mostly declarative ('And then E and an N and an E' (Expert B, III 663) 'And then F E C I T' (Expert A, III 663)). This can be seen by comparing the average time spent discussing the different reading levels with the amount of separate instances of these discussions which occur during the transcripts.

The subjects which are discussed for the longest length of time per instance are the physical characteristics of the document, and the overall meaning of the text. The information regarding these levels is much more complex than the identification of characters and words, and tends to be discursive. For example, when discussing the meaning of ink tablet III 616 Expert C explains:

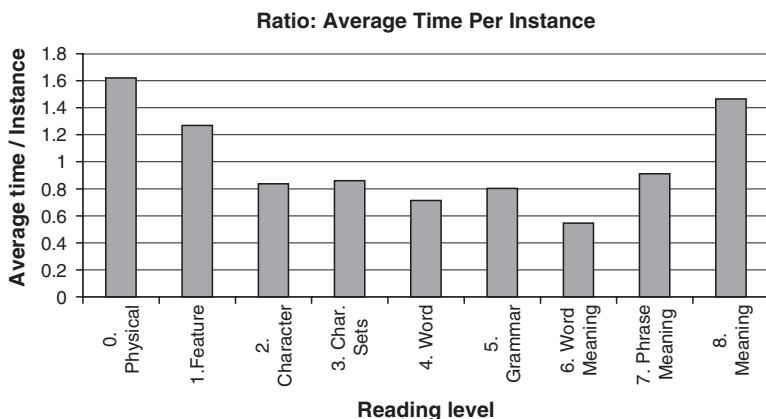


Figure 2.9. The average time spent discussing each subject compared to the amount of separate occurrences of those discussions within the transcripts. A ratio of 1 indicates the same length of time as the instances of discussion. Over 1 indicates that discussions are longer than average per instance, less than 1 indicates discussions are shorter than average.

This all seems to make good sense. An instruction . . . will all of you diligently take care that if any friends come they are well received, or something like that. He seems to have got a clear sense, starting here, which would fit very well with the beginning of a letter or instruction, which suggests that we probably have got the top of the tablet, and that this is a letter of instruction, possibly to the freed men, possibly even to the slaves as well . . .

Discussions regarding the features which constitute characters, however, tend to be both frequent, and lengthy:

And then, I'm afraid, more of these wretched things, this unit is part of so many different letters, its basically part of an A but could be part of an M, can even be part of a rather slanting N, I've got sort of three of these things, here. It's really rather difficult to tell them apart. (Expert B, 797)

This indicates that the feature level is one of the most key, and most complex, in the process. It can be taken from this that the identification and rationalization of what may and may not be a part of a letter is one of the most trying stages in the reading of ancient texts.

A summary of the frequencies of topic, and the length and type of discussions regarding each topic can be found in Table 2.7.

2.5.4. Reading Ink Texts Versus Reading Stylus Tablets

An analysis of the Think Aloud Protocols also indicates the differences between reading the ink texts (carbon ink on wood) and the stylus texts (incised texts) (see S. 1.1). Although from the same period and using similar letter forms (see S. 3.1.1), the stylus tablets are much harder to read due to their physical characteristics. Exactly how this affects the discussions regarding the texts can be seen in Figure 2.10.

It can be clearly seen that there are more instances of the feature and character level in discussions of the stylus tablets rather than the ink texts. There are more instances of the word and meaning levels in the discussions regarding the ink texts, indicating how hard it is to generate meaning from the writing on the stylus tablets: there are few words identified in the stylus texts to talk about! This is backed up when comparing the length of time different subjects are discussed in the reading of the ink and stylus texts.

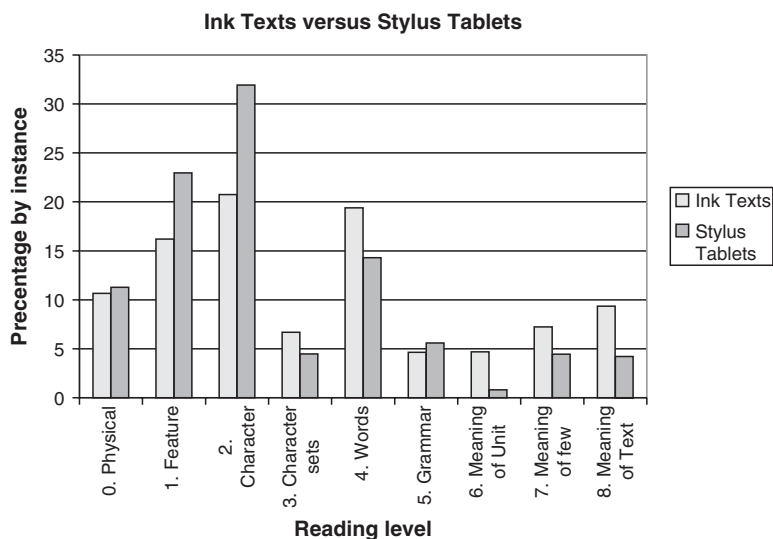


Figure 2.10. Comparison of percentages of subject matter in discussions of the ink and stylus texts. Two ink texts and two stylus texts discussed by Expert A and Expert B were used to generate the data.

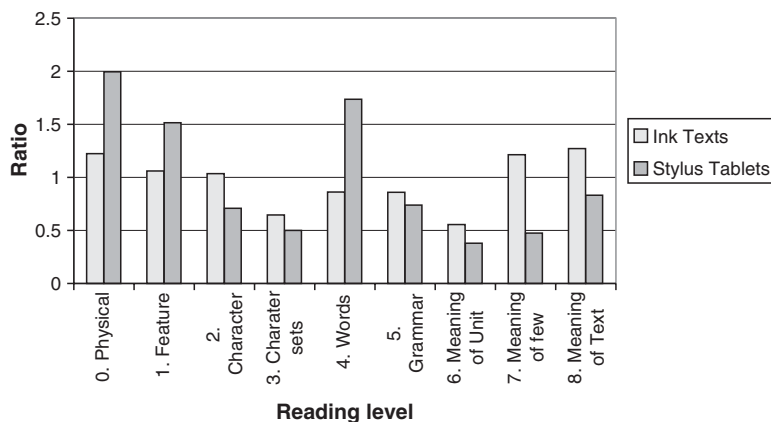


Figure 2.11. The average time spent discussing each subject compared to the amount of separate occurrences of those discussions within the transcripts of the ink and stylus texts. A ratio of 1 indicates the same length of time as the instances of discussion. Over 1 indicates that discussions are longer than average per instance, less than 1 indicates discussions are shorter than average.

Discussions regarding features of characters in the stylus texts are lengthy as well as frequent, as above in section 2.5.2, and are much more lengthy than those of the ink texts. The discussion of features is therefore crucial to the reading of stylus texts.²³ Interestingly, the discussions regarding the physical attributes of stylus texts are lengthy and complex, indicating that much more attention is paid to the format of the text when reading stylus texts. Although discussions regarding words are not as frequent in the stylus texts as in the ink texts (Figure 2.10), discussions identifying words in the stylus texts are much lengthier, indicating how hard it is to identify words in these type of documents. Table 2.7, which summarizes the different characteristics of these writing levels, is presented in section 2.7.

²³ It has been noted that ‘the more difficulty the reader has with reading [a text], the more he relies on the visual information’ (Smith 1971: 221).

The difficulty in identifying features, characters, and words in the stylus texts can also be illustrated by comparing the word lists derived from the ink and stylus text discussions. The key words in the discussion of the stylus texts (which are unusually frequent in comparison with what one would expect on the basis of the discussions regarding ink texts) are shown in Table 2.6.

Although these words are distinctly unglamorous, they show that the experts need to continually point out elements of the text (here, there, this), try to evaluate their judgements more (sort, whether this is used to 'sort' between hypothesis, or explain how they 'sort of' function), and explain to the knowledge engineer the features that they see, and how they undertake the task (you).

A further comparison of the word lists derived from the two sets of discussions illustrates the difference in approach towards the two types of texts further: the experts use the words AFRAID, ASSUME, CONFUSING, CONVINCE, DECIDING, SURPRISED, and TRIED in relation to the stylus texts but not the ink texts. Words occurring in the ink text discussions but not the stylus text discussions tend to be less about the decision-making process and the frame of mind with which the experts approach the texts, and more about features of the texts themselves, such as HORIZONTAL, BOLD, FORMAT, and DISCOLORATION.

Reading the stylus texts is more laborious than reading the ink texts, with much more attention to the features of characters necessary, and much more ambiguity surrounding the hypothesis generated regarding word identification. Less attention is given to discussing the meaning of the document as a whole, mainly because they do not seem to get that far in these two examples: the experts have problems enough in making certain assumptions about the

Table 2.6. Key words in the discussions regarding stylus tablets.

N	WORD	FREQ.	STYLUS TXT %	FREQ.	INK TXT %	KEYNESS
1	HERE	100	1.14	30	0.21	83.8
2	THERE	224	2.56	182	1.26	52.1
3	THIS	195	2.23	170	1.17	38
4	SORT	83	0.95	53	0.37	30.6
5	YOU	173	1.98	163	1.12	26.9

Keyness is calculated using the classic chi-square test of significance with Yates correction; any word with a keyness of 25 or over is taken to be significant (Sinclair 1991).

characters and words within the texts. Further investigation into this area could look to see if there was any difference in the order that experts discussed different types of knowledge between the stylus and ink texts, the data sets here being rather small to allow valid conclusions to be drawn.

2.5.5. Recounting the Process

There are two further questions which shall be addressed regarding the process of ‘private’ papyrology. First, how does the initial reading of a text relate to an explanation of that reading at a later date (and so how valid is the evidence presented by papyrologists when they explain how they reached a reading)? Second, how does the subject and textual content of the expert’s discussions regarding the texts relate when compared to those of the published volume of apparatus containing the ink texts, *Tabulae Vindolandenses II*?

2.5.5.1. Retelling the story

It has already been shown that when an expert was asked to recount how he came to an understanding of a text after an initial reading, he discussed the prime elements in a different order (Figure 2.6). The primary tenet of knowledge elicitation is that experts find it difficult to relay the processes they went through whilst undertaking a task. However, analysis of concurrent and retrospective Think Aloud Protocols by Ericsson and Simon (1993) shows that there is very little difference between the two. Expert B’s two discussions regarding III 663 are very similar. The first discussion lasts 1,347 seconds, the second 1,246 seconds, and the frequency of subject matter discussed correlates closely, as shown in Figure 2.12.

When an error rate of 4.1 per cent is considered, it becomes obvious that there is very little difference in the distribution of subject matter between the two readings. The same is also true for the length of time talking about the subject matter in each case. There is also little difference between the words used to discuss the texts, and the frequencies or collocates of these words. It would seem that the retrospective discussions of how the experts approach a text are

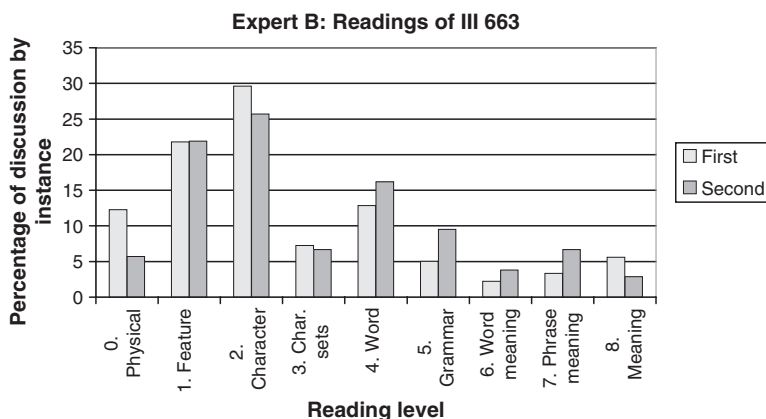


Figure 2.12. Comparison of Expert B's reading and retrospective account of reading III 663.

closely linked to the actual processes they got through (or the ones they think that they go through and can verbalize) when trying to read a text. The problem for the knowledge engineer, then, is not when the experts discuss how they carried out a process, but if there is any difference between their account and what is actually taking place, a gulf there is no means to cross at present.

2.5.5.2. *Published commentaries versus discussions*

How do the published commentaries regarding the ink texts differ from the discussions regarding the reading of the ink texts, and what can they show about the process of reading an ink text? A content analysis of seven critical apparatus from *Tabulae Vindolandenses II* was compared with the results from a similar analysis of the discussions regarding the seven ink texts, to see the difference in subject matter covered between the two (Figure 2.13).

It is obvious that the published commentaries focus mostly on the identification of words, whilst the Think Aloud Protocols concentrate equally on the identification of features, characters, and words, and place attention on the physical nature of the document. This is partly due to the publishing format, and aims, of the apparatus, but

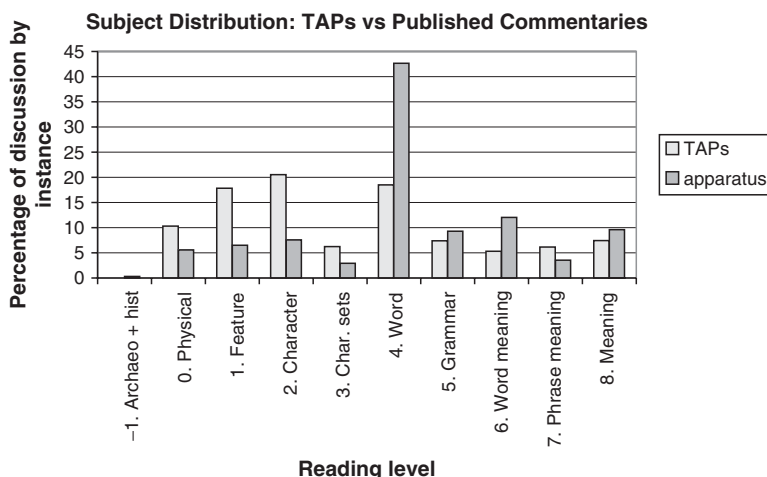


Figure 2.13. Distribution of subject matter in the Think Aloud Protocols compared with those of the critical apparatus of the ink texts.

also shows that the published commentaries are slightly removed from representing how texts are read, not describing the basic process of reading a text. However, the Leiden system does preserve some information regarding the certainties of the readings of the texts, and an analysis of this provides some interesting information about the role of certainty in the reading of ancient documents.

2.5.5.3. *Textual analysis of the published Vindolanda ink texts*

The Leiden system of transcription, a series of symbols that denote various characteristics of the original text,²⁴ was devised in 1931 at a meeting of papyrologists in Leiden (van Groningen 1932; Turner 1980) and remains the standard in notating the transcription of a document.²⁵

²⁴ The Leiden system is as follows: [] probable text which is missing from the document; () an abbreviation which has been expanded; { } an editorial deletion; [[]] an original scribal correction or deletion; ‘ ’ an insertion above the line; (below a letter) a letter which cannot be read with certainty.

²⁵ Although the application of the Leiden system can be shown to vary between different readings of a text by both the same and separate individuals (Youtie 1966).

The published Vindolanda ink texts annotated in this manner from *Tabulae Vindolandenses II* (Bowman and Thomas 1994) and the ink texts that had been prepared for publication for *Tabulae Vindolandenses III* (Bowman and Thomas 2003) were collated. This gave a corpus of 286 ink texts in total,²⁶ comprising 2,301 lines and 27,364 characters of Vulgar Latin, grouped into 6,532 words, or groupings of characters.

Of these 27,364 characters, 25,915 characters were identified as actually being present within the documents, the remaining 1,449 being probable text which is missing from the document due to breakage, damage, etc. These were grouped in 761 separate instances, often ending or beginning words which were left incomplete due to damage. This means that 5.3 per cent of the characters, and 11.7 per cent of the words in the readings of the Vindolanda texts were actually not physically present, and predicted by the experts because of their prior knowledge of the language and similar texts, being restoration rather than reading. There were 772 (2.8%) instances of characters taken to be abbreviations within the documents.

From the total 25,915 characters, 2,566 were under-dotted, indicating an uncertainty about the reading of the letter. There is no way to measure the extent of that uncertainty (i.e. whether they are a little uncertain of the reading, or very unsure that it is the correct reading), practice varies between papyrologists, with some using the underdot to represent a broken letter which is certain, others using it only when identification is in doubt. The authors of the volume tend to use it both when identification is uncertain and when there is no doubt: using the dot to show that the letter is broken. Therefore, 9.9 per cent of the characters read in the Vindolanda ink texts were marked as being broken: a fairly high proportion, and the identification of an unknown proportion of these will be problematic.

Thus ambiguity is a main feature of the readings reached of the documents, even in their published version. This is not a critique of the work of the papyrologists; published versions are open to correction, and merely detail the extent to which the author has

²⁶ Although over a thousand tablets in total have been found at Vindolanda so far, many of them contain little or no writing, or are in such a fragmentary state that nothing of value can be read from them (Thomas 1992), and so only the published texts were included in this corpus.

resolved the reading of the text at that moment in time.²⁷ However, it does show that there still remains a great deal of uncertainty about the reading of the ink texts, and that this is very seldom exhausted. The process revolves around the resolution of ambiguity through prediction, prior knowledge, reasoning about the characteristics of the document, and the addressing of uncertainties. This ambiguity would be difficult to implement in an artificial intelligence system, as it would be hard to provide enough real-world knowledge to reflect the amount of contextual information necessary to undertake such a task.

2.6. GENERAL OBSERVATIONS

There was substantial procedural information gathered from the literature review, the elementary stages of knowledge elicitation, and the data from the Think Aloud Protocols. These general observations regarding the reading of the ink and stylus tablets illustrate further the complexity of the reading process. Some observations peculiar to the reading of stylus tablets indicate the difficulties faced by the experts when trying to read such damaged and abraded texts.

The papyrologists continually refer to other material, such as other texts, word lists, grammars, and archaeological evidence.²⁸ Reference to such information is not always made explicit. Only when all parallels to other texts are exhausted are they comfortable with making any new assertions regarding the language or subject of a document: there is an element of appropriateness to the resolution of ambiguity regarding the texts.

²⁷ Bowman and Tomlin (2005) provide examples of readings which have changed dramatically between different versions of published texts. Tomlin (1994) demonstrates how the correct reading of a lead tablet was achieved by rotating the tablet through 180 degrees and rereading the text.

²⁸ The importance of contextual information—understanding other documents, and the archaeological and historical evidence of the time, as well as the linguistic and palaeographic data which is likely to be contained in the texts—is common to those trying to ascertain the readings of any type of document. See Clanchy (1993: 82) for the techniques used when trying to read medieval texts, and Munby *et al.* (2002) for techniques used to read Tudor and Stuart Handwriting.

The understanding of convention is paramount in being able to read a text. There are both physical conventions, regarding format, and textual conventions, regarding language. The format of the document can indicate whether it is a letter, an account, or official documentation, and these standard formats can allow the experts to immediately focus expectations regarding the possible subject matter of the texts. Linguistically, there exist conventions of letter writing, phrases commonly used to express certain situations, and conventional ways of conveying certain information, for example, literary phrases and high-class spellings. The papyrologist is always on the look out for any conventions at work (and alternatively, unconventional usage) in order to gain an understanding of the text. As the body of texts read from Vindolanda increases, the conventions regarding format and language are becoming better understood, and so aid in the reading of subsequent documents.

The starting point in reading a document is always the clearest text on the document, not the first characters at the beginning of the document. Uninscribed or clear areas (whether or not deliberate word separation has been used) can give vital first clues to where words begin and end, and the space between these can indicate the number of characters needed to complete the lexical unit. The papyrologists often note the relationship of the writing with the grain of the wood and how it flows along the line, which can give some indication of the letter forms. There is also consideration of the physical act of writing, asking what could be the possible intention of the writer, and if there was a problem in making individual characters due to the format, for example, squeezing in characters so that a word fits into the end of a line. Differentiation in the thickness of strokes in the ink tablets can help identify characters.

Once a letter form has been identified in the document, this is often used as a template to compare more ambiguous letters, to give some measure of the likelihood that a group of features can be identified as a letter. The expert's familiarity with the different handwriting already identified in the corpus can give some indication of the subject matter of the text; if they can identify the writer they can have a good idea of who the document was addressed to, and the sort of information it will probably contain.

Probable sequences of letters give a clue to what words might be there, and it is often easier to work backwards from the end of word than the beginning to identify it. Possible words are continually checked with external sources. Deciding whether the words wrap around to the next line of text can prove problematic. The identification of words does not happen in isolation: context is everything.²⁹ It is perfectly acceptable to leave readings as conjecture when they still remain unclear.

As far as the stylus texts are concerned, the first task is deciding which way up the document goes, which is often tricky.³⁰ The format is paid more attention to, as it can often give a clearer indication as to the purpose of the document than the format of the ink texts, which is more uniform. It can be difficult to decide whether the writing on the stylus text continues all the way across the document, or if it is written in columns. Possible lines of text are identified, to give an indication of where characters may actually be. Uninscribed areas are particularly important in the reading of the stylus texts, as they provide some definite information regarding the flow of text.

The inscribed nature of the writing with the stylus point means a signature mark is left when it ‘bumps’ over woodgrain, and this can differentiate actual letter strokes from background noise. Care needs to be taken to accommodate the effect of changes of pressure in the making of strokes in order to identify them properly, as tails of letters can often lift off and become faint. Strong definite strokes are identified first, and signature strokes, such as descenders and ascenders, can give clues as the use of key letters, for example, the letter S (see S. 3.7). Care must be taken to avoid reading persistent strokes that remain from text(s) written on the stylus tablet previously. Some consideration must be given to differences in letter forms (from those used in the ink texts) caused by writing on a different, and more difficult, medium (see S. 3.1.2).

²⁹ This may present problems when moving to a more fully automated system, in that trying to model the contextual information is almost impossible due to the vast nature of associated information that papyrologists use when trying to read a text.

³⁰ ‘I decided it was probably the other way up I think, did I or did I not? No I think it is this way up’ (Expert B, 974).

2.7. PRELIMINARY CONCLUSIONS

Although each expert has his own individual style in reading an ancient text, it has been shown that there are some unifying procedures in the process of reading an ancient document. The first success of the knowledge elicitation exercise was making explicit the different topics discussed regarding the ink and stylus tablets, providing a representation of the different types of information the experts use whilst reasoning about such texts. The research also enabled the identification of characteristics regarding each level, and the frequency and order of occurrence of each topic. Moreover, this gave specific information regarding how reading the stylus tablets differs from reading the ink texts, indicating the features of reading that become more important when encountering more damaged and abraded texts. This information has been summarized in Table 2.7.

The investigation also revealed the cyclic reading process; the process of reading a text is not linear, building up one character at

Table 2.7. Features of the reading levels identified in the discussions regarding the ink and stylus texts.

Reading level	Thematic subject	Characteristics
8	Meaning or sense of document as a whole	This is discussed surprisingly rarely, but when it is, the reasoning tends to be lengthy, complex, and discursive. It is usually preceded by discussions regarding the identification of words, and followed by discussions regarding the physical characteristics of the document, features of characters, and the identification of characters. There were more often, and lengthier, discussions regarding this level in relation to the ink texts than the stylus tablets.
7	Meaning or sense of a group or phrase or words	This is rarely discussed, and tends to be regarding linguistic convention. It is usually preceded by identification of words, and followed by discussions regarding physical characteristics and the identification of character sets. Discussions tend to be short, although they are marginally longer with regard to the ink texts. This level occurs marginally more often with regard to the ink texts.

(continued)

Table 2.7. (*Continued*)

Reading level	Thematic subject	Characteristics
6	Meaning or sense of a word	Surprisingly, this is seldom discussed. It is the shortest of all discussions, tending to be very declarative. It is usually preceded by the identification of a word, and followed by discussions regarding grammar. There are marginally more occurrences of this level with regard to the ink texts.
5	Discussion of grammar	This is seldom discussed (although Expert C refers to grammar often). It is usually preceded by the identification of a word, and followed by discussions regarding the feature level. Discussions are short, and are distributed equally over the transcripts regarding the ink and stylus texts.
4	Identification of possible word or morphemic unit	One of the most commonly discussed subjects by the papyrologists, although discussions tend to be short and declarative. It tends to be preceded by the feature, character, character set, and word levels, and is often followed by every other level in the representation: the only one to have this characteristic. Discussions regarding words are more frequent in relation to the ink texts, but are considerably longer in relation to the stylus texts.
3	Identification of sequence of characters	This is fairly seldom discussed, and discussions tend to be short. It is usually preceded by discussions regarding the feature and character levels, and followed by discussions regarding the identification of characters and words. It is equally distributed in both sets of transcripts.
2	Identification of possible character	This is the most discussed topic, on average. Discussions tend to be declarative and short. It can be preceded by discussions regarding physical characteristics, features, characters, and less often discussions regarding character sets, phrase meaning, and the meaning of documents. It is most often followed by discussions regarding features, characters, character sets, words, and sometimes the physical characteristics of the document. It is much more frequent in the transcripts regarding the stylus tablets, but discussions regarding this in relation to the stylus tablets tend to be shorter.
1	Discussion of features of character	This is one of the most commonly talked about topics in the transcripts, but discussions tend to be complex, lengthy, and discursive. It is usually preceded by discussions regarding the physical characteristics of the document, the feature of characters, the identification of characters, and sometimes identification of words and character sets. It is usually followed by discussions regarding features, characters, character sets, and identification of words. It is more frequent in relation to the stylus texts, and discussions of this level also tend to be lengthier in relation to the stylus texts.

0	Discussion of physical attributes of the document	This is discussed fairly regularly, and discussions tend to be lengthy, complex, and discursive. It is often followed by discussions regarding the physical attributes, features, and identification of characters, and often preceded by discussions regarding the physical characteristics, features, and characters. It is important in both sets of transcripts, but discussions of this level in relation to the stylus tablets tend to be lengthier.
-1	Archaeological or historical context	Direct reference to external resources was surprisingly rare. However, the experts continually relate the hypotheses they generate to their own extensive knowledge regarding similar texts. This level was mostly present in the published commentaries, providing examples to aid in the discussions of other levels.

a time, but depends on the propagation of hypotheses, and the testing of these regarding all available information concerning a text. Reading a document is a process of resolution of ambiguity, and depends on the interaction of all the different facets of knowledge available to the expert. Specific details of how the experts proceed in reading ancient documents were also made explicit.

These conclusions can be drawn together to propose an elementary model of how experts read ancient texts. (However, first it will be necessary to review other models of reading that have been developed in the field of psychology.) The conclusions can also be used to indicate which part of the process the experts need assistance with to aid in the reading of the remainder of the stylus texts. The lower levels of the process (discussions regarding physical attributes, identification of features, identification of characters, and words) are much more a focus in discussions regarding the stylus tablets, and any subsequent computational tools should concentrate on aiding the papyrologists in these areas. This is the focus of the system discussed in Chapter 4 and Appendix A.

2.8. MODELS OF READING AND PAPYROLOGY

Although the act of reading an ancient document has never been the focus of a psychological investigation, and there exists very little psychological research that directly relates to the reading of an

ancient text (see S. 2.1.2), there exists a large body of research covering general aspects of reading. Various attempts have been made to construct some kind of model of the reading process, but it has so far proved impossible to condense all the different elements involved into one precise, albeit high-level, theory. However, successful models exist which relate to specific tasks in reading. Using the results from the investigation, above, it is possible to propose a model of how experts approach and read ancient texts.

2.8.1. Psychology and Models of Reading

Reading, as a focus of cognitive research, covers a large, problematic, critically diverse area. Constructing models of the reading process has proved problematic, but so has defining what 'to read' actually means; 'Reading is extracting information from text.' (Gibson and Levin 1976: 5); 'Readers do not so much extract meaning from print, but rather engage in an active construction of meaning based on the signs provided by the print' (Crain and Steedman 1985: 321). Although the physical process of reading is well documented, for example, the physiology of the eye and the muscle movements made whilst reading a text (Gibson and Levin 1976; Oakhill and Garnham 1988; Asher 1994; Gregory 1994; Manguel 1997), the cognitive process that results in assigning a text meaning remains obscure.

Many attempts have been made to illustrate the reading process using systems modelling techniques borrowed from the field of computer science. Two types of these models of reading exist: sequential and componential. Componential models, which aim to break the task of reading into separate distinct components, such as decoding, language comprehension, and fluency (Carr and Levy 1990), have proven problematic as it is impossible to form a definitive list of universally applicable components of the reading process (Asher 1994). Sequential models have been slightly more successful in their aim to chart the time course of reading from the start of the process, and can be classified into three sections. 'Bottom-up' models such as Gough's 'One Second of Reading' (Gough 1972; Gough and Cosky 1977) utilize perceptual information and skills, treating each occurrence of reading as a new cognitive puzzle to be

solved individually. 'Top-Down' models, such as Goodman's 'psycholinguistic guessing game' (1967) define reading as a process of prediction, confirmation, and correction and rely on previously stored linguistic information. Interactive, or connectionist, models, such as McClelland and Rumelhart's 'Interactive Activation and Competition Model of Word Recognition' (McClelland and Rumelhart 1981) cycle between the top-down and bottom-up processes, and are the most favoured of all models by current theorists (Ellis and Humphreys 1999).

The extent to which these models actually explain the reading process can be illustrated by sections of Gough's model: 'Merlin' is the name given to the mechanism that magically applies syntactic and semantic rules to viewed text, and the central processing plant is named TPWSGWTAU: 'The Place Where Sentences Go When They Are Understood'. In recent years attempts to develop a definitive model of the reading process has slowed as it has become obvious that reading is a complex system which encompasses various diverse processes: 'all that can be said with a fair degree of certainty is that the skills readers use for the extraction of meaning from print are extremely complex, automated to a fairly high degree, and highly flexible, comprising a number of different styles' (Asher 1994: 3457).

The flexibility of the reading process has led researchers to identify a number of reading styles, dependent on both reading purpose and text type, for example, skim reading, detailed reading, and puzzle solving (Tunmer and Hoover 1992; Asher 1994; Gibson and Levin 1976). As 'there is no single reading process, there can be no single model for reading' (Gibson and Levin 1976: 438). However, they can give insight into individual processes, such as McClelland and Rumelhart's 'Interactive Activation and Competition Model of Word Recognition' (McClelland and Rumelhart 1981, 1982, 1986*b*; Rumelhart and McClelland 1982) which aims to explain the mechanism of the word superiority effect (Figure 2.14).

2.8.2. The Interactive Activation Model of Reading

That a word can be identified more accurately than a single letter has been known since the earliest considerations of the reading process

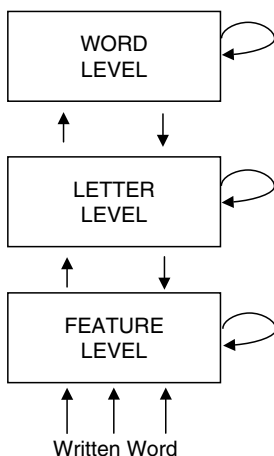


Figure 2.14. McClelland and Rumelhart's 'Interactive Activation and Competition Model of Word Recognition' (1981: 379; see also Rumelhart and McClelland 1982, 1986; McClelland and Rumelhart 1986*a*, 1986*b*).

(Cattell 1886). The word superiority effect, the fact that 'it is possible to identify letters in words more accurately than letters in random sequences' (McClelland and Rumelhart 1981: 376) is accepted by most in the field.

McClelland and Rumelhart's model demonstrates the resolution of ambiguity in the reading process by the use of a recursive mechanism, which favours the identification of letters in known words. Features are identified at the basic level and then checked to see if they are likely characters. If not, the features are reidentified. Once a letter appears to be likely, it is related to the surrounding letters to see if they are likely to combine and become a word. Words which are already known are therefore most likely to be the result of this process, and words take precedence over identifying random characters independently: a character is most likely to be identified if it is part of a known word. This is a very simple, recursive, model, shown to be very effective in practice (Ellis and Humphreys 1999). An implementation of a model similar to this was used by Robertson (2001) to demonstrate the effectiveness of the architecture of a system based on minimum description length (see Chapter 4 and

Appendix A). Robertson's system, based on this recursive model, successfully 'reads' a handwritten text.

2.8.3. Proposed Model of the Papyrology Process

The papyrologists seem to operate in a similar way as the model shown above, as it has been demonstrated that they use a recursive reading mechanism which oscillates between different levels, or modules, of reading. To this extent, it is possible to propose an overall model of how the experts work. The model shown in Figure 2.15 is stacked hierarchically, with the 'overall meaning of text' agent nominally being taken to be the highest level in the sequence of

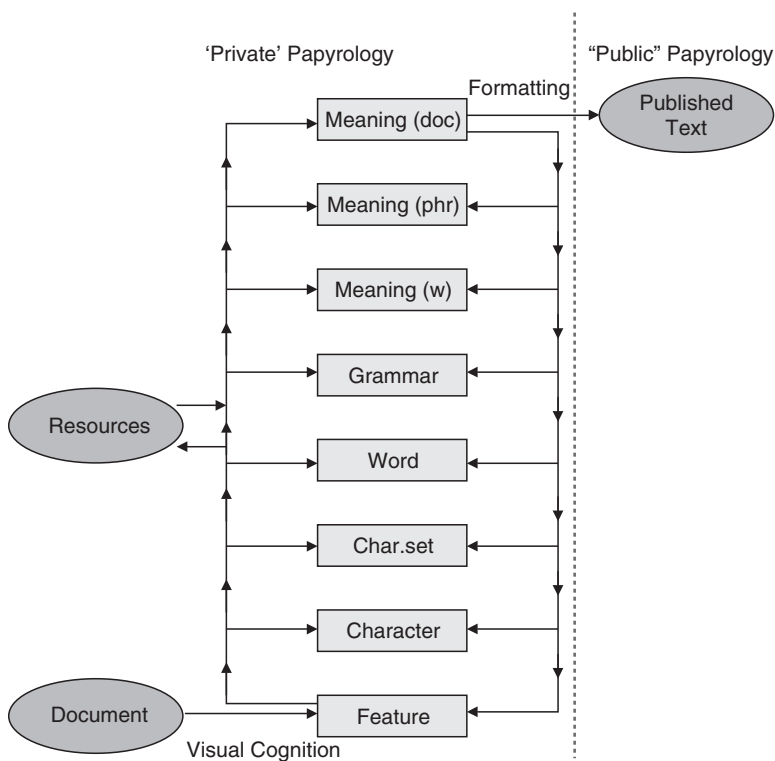


Figure 2.15. Proposed model of the procedure used to read an ancient text.

events before a transcription of the text can be prepared for publication. As suggested above, it is possible that the word level is actually the most important of the process, but the levels are presented in this order to suggest a basic order of interaction whilst trying to develop a reading of the text.

An expert reads an ancient document by identifying visual features, and then incrementally building up knowledge about the document's characters, words, grammar, phrases, and meaning, continually proposing hypotheses, and checking those against other information, until s/he finds that this process is exhausted. At this point a representation of the text is prepared in the standard publication format. At each level, external resources may be consulted, or be unconsciously compared to the characteristics of the document.

Although this is a simple representation of the process undertaken, it shows the overall scope of the process of reading an ancient text. This model was used as the basis of the computer system developed to aid in the reading of the stylus texts, as discussed in Chapter 4 and Appendix A.³¹

2.9. CONCLUSION

This investigation has been the first attempt to make explicit the process of reading an ancient text. Although it has been shown that individual experts have different approaches to reading a text, it has been demonstrated that there are overall similarities in their working methods. The process has been stratified into defined modules, investigating the characteristics of each, and how they relate to each other. The problems in reading ink and stylus texts have been illuminated, with specific conclusions drawn regarding how reading the stylus texts differs from reading the ink texts. The role of ambiguity in the production of readings has been highlighted. The general process has been condensed into an overall model in order to provide a basis for the development of computational tools to aid the experts read the stylus tablets.

³¹ Details of the relationship between the model and the system can also be found in Robertson *et al.* (forthcoming), and Terras and Robertson (2005).

There is no doubt that the organization, documentation, and analysis of the Think Aloud Protocols and published commentaries was a time-consuming and tedious task. Some of the findings regarding the papyrology process may seem obvious (the more difficult and deteriorated a document, the more the experts pay attention to features of the document: who would have thought?). However, it has provided data which have successfully given insight into a hitherto undocumented area, and has shown how utilizing such techniques can provide rich quantitative and qualitative evidence on how experts carry out their tasks. There remains a great amount of research which could be done to investigate the cognitive processes the experts go through when reading an ancient text. For example, more graded tasks could be given to the experts. The role of eye movements in reading such texts could be investigated; little attention has been paid to the role of visual cognition in reading ancient texts. The techniques of experts who work with other language systems could be studied. Possible future work regarding knowledge elicitation is detailed in Chapter 6.

The conclusions reached in this chapter also pertain to the process of reading texts of a particular period, location, and medium, and it is more than possible that an investigation into the reading of different types of ancient document and inscriptions may reveal different findings (although it is suspected that there will be similarities between reading different writing systems). However, this research has provided a concrete starting point for the development of computational tools to aid papyrologists in reading the Vindolanda texts, by providing a model on which to base the development of a computer system.

The Palaeography of Vindolanda

A language is a dialect with an army and a navy.

(Weinreich 1945: 13)

To enable the construction of a system which can be used to read the Vindolanda stylus tablets, it was imperative to gain an understanding of the letter forms and language that might be contained within the texts. The major significant body of contemporaneous documents to the stylus tablets are the Vindolanda ink texts, and it can be demonstrated that the stylus tablets should contain similar character forms to those found on the ink tablets, even though the two types of texts are written on substantially different media. Data were needed to train a system to read unknown characters contained within the stylus tablets. The construction of a corpus of annotated images, based on the letter forms found in the ink tablets (and the few stylus tablets that have been read), and the gathering of linguistic information from the Vindolanda ink texts that have already been read, provided the information needed to train the system. This chapter focuses on the methodologies used to collate these data sets.

First, it was necessary to undertake a review of what is known about the letter forms contained in the Vindolanda texts. Both the ink and stylus tablets contain handwriting in the form that is known as Old Roman Cursive (ORC). Although many aspects of Latin palaeography have been studied in detail, ORC is the focus of academic debate, due to the paucity of documents available from this period. A series of knowledge elicitation exercises were undertaken with the three experts (as consulted in Chapter 2) to gain an understanding of the types of information they use to describe and identify ORC character forms. From this a schema was constructed, detailing

the relationships between the different types of information identified. An encoding scheme was developed from this schema, enabling the annotation of images of the Vindolanda texts, producing textual descriptions of the letter forms in Extensible Markup Language (XML) (extending the focus and function of the marking up of text in the humanities, as most textual encoding in the arts has involved marking up literary and linguistic digital texts from the character level upwards, rather than annotating images of the texts to encapsulate information about the stroke detail which constitutes the text in question).

Nine documents were annotated, using this encoding scheme and an annotation program which was adapted from its original use of capturing data regarding aerial images. This resulted in a corpus of images annotated in XML data comprising of 1,700 individual letters from the Vindolanda corpus. Additional linguistic analysis of the Vindolanda ink texts generated lexicostatistics (word lists, word frequency, and letter frequency) which also provide much data regarding the language used at Vindolanda. As such, the data sets constitute a unique information source that can be used by palaeographers and papyrologists, and are the source of information used to train the system described in Chapters 4 and 5, and Appendix A. The corpus can also be used to undertake a comprehensive comparison of the differences between Old Roman Cursive as found on the Vindolanda stylus tablets and the ink texts, and it is demonstrated here that the letter forms found within the ink and stylus tablets are indeed similar (backing up the earlier assumption about the similarity of the letter forms on the two types of documents).

3.1. PALAEOGRAPHY

Palaeography, the study of ancient handwriting, ‘in the strictest sense deals only with the old styles of writing, whereas palaeography in the wider sense embraces everything related to written texts of the past: the technique of writing, its material support, the mode of its transcription, and also the ways texts were diffused and circulated’ (Boyle 2001: p. xi). Palaeography incorporates elements of codicology,

epigraphy, philology, diplomacy, and papyrology. Its traditional task was to date manuscripts, and also to investigate the authenticity of documents (Bately *et al.* 1993). Its methodology is fairly transparent, as the palaeographer deals with the forms of letters on an individual basis, providing in-depth documentation¹ to chart the development and use of styles of handwriting. The handwriting of Latin manuscripts has been the focus of systematic research since the late seventeenth century (see Bischoff (1990) for a comprehensive introduction).

3.1.1. The Palaeography of the Vindolanda Ink Tablets

The letter forms in the Vindolanda ink tablets are ‘Old Roman Cursive’ (ORC)² (Bowman and Thomas 1983), the everyday Roman script during the first three centuries AD (as opposed to the formal script used for literary works). ORC is characterized by its small, unelaborate characters: differentiating it from the capital forms used in the ‘literary bookhand’ (Bowman 2003). Although ORC was the commonly used hand, and has been studied since the early 1800s, extant sources are rare,³ and are mostly from the south and eastern reaches of the empire. Scholars’ understanding of the forms and conventions utilized is still incomplete, despite much academic interest in the area.⁴

¹ e.g. a comprehensive chart of letter forms from the Bath Curse Tablets, and discussions regarding these characters, can be found in Tomlin (1988).

² This form of writing has also been called ‘Scrittura usuale’ (Cencetti 1950), ‘l’écriture commune classique’ (Mallon 1952), ‘Ancient Latin Cursive’ (Thomas 1976), and ‘Ancient Roman Cursive’ (Tjäder 1986). It will be called ORC here for the sake of consistency.

³ The primary sources for ORC, aside from the Vindolanda corpus, are 200 tablets from Pompeii and Herculaneum (pre-AD 79), tablets from Dacia (AD 131–67), a variety of tablets from Egypt (1st–4th centuries AD), 400 tablets from Vindonissa in Switzerland (mid-1st century AD), and a handful of tablets from Britain. These sources are discussed in Bowman and Thomas (1983: 33–7). A complete list of known examples up to 1955 was compiled by Marichal (1950, 1955).

⁴ For a comprehensive bibliography, see Bowman and Thomas (1994), and Tjäder (1977).

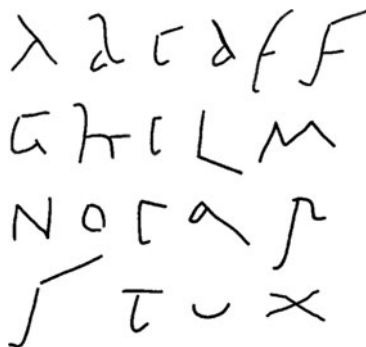


Figure 3.1. Characteristic letter forms in ORC, based on Bowman and Thomas (1983: 54). The alphabet consists of 20 characters (not having the letters J, K, V, W, Y, and Z present in modern-day English script).

From around AD 300, ORC was replaced by New Roman Cursive (NRC), often referred to as ‘miniscule cursive’, the transition between the two being the focus of much academic debate (see Tjäder 1979). (NRC led indirectly to Carolingian miniscule in about 800 AD, and so to the script we employ today.) The documents from Vindolanda are from a much earlier period than that of NRC, and the ink and stylus tablets provide an excellent palaeographical source for researching the development of ORC Latin script,⁵ and the use of ORC in Roman Britain. Bowman and Thomas explain:

only a few of the tablets will actually have been written at Vindolanda, but the remainder are hardly likely... to have travelled far, and we may fairly take the find as exemplifying the type of writing in use in Britain at this period. We thus have now for the first time a not inconsiderable body of written material from a part of the Empire from which hitherto virtually no such material had come to light. (1983: 52)

Bowman and Thomas provide a comprehensive discussion of the individual letter forms and other general palaeographic detail regarding the ink tablets (1983, 1994).

⁵ There are other sources of ORC script which this research could be extended to include, such as other material from Pompeii and Herculaneum: see Bowman and Thomas (1983: 33–7) for a bibliography. However, the letter forms contained on the Vindolanda ink texts are the closest and most obvious point of contact to the letter forms which may appear on the stylus texts—and the most readily available.

3.1.2. The Palaeography of the Stylus Tablets

In order to aid in the reading of the stylus tablets, it is important to know what type of letter forms they will contain. However, due to the fact that only a handful of the Vindolanda stylus tablets have been read so far, it is impossible to provide a traditional palaeographic review of all the letter forms used within them. Those that have been read utilize ORC script, having the same letter forms as the ink tablets. This is fairly unsurprising as they are contemporaneous documents from the same source, although their content is expected to be more legal in nature (see S. 1.2). There has been some discussion of different forms of letters being employed in official documents (Bischoff 1990), but this has tended to apply to comparisons involving papyri and stylus tablets, and there is little evidence that this was the case in Vindolanda. Cencetti maintains that around this period the script in legal documents as well as in military and civilian administration took on a very uniform character (1950). The many individual hands (over three hundred) present in the ink tablets, which are both official and civilian correspondence, utilize the same script, showing ‘that there was a single, standard type of script in common use, and that this script could be written both quickly and slowly’ (Thomas 1976: 41). There is no reason to think that the letter forms in the Vindolanda stylus tablets will differ wildly from those in the ink tablets, aside for individual cases, such as the letter E, which is written in a different format on wax in most cases.



Figure 3.2. Letter E. On the left is the form commonly found on the ink tablets. On the right is the form commonly found on stylus tablets. The ink form is occasionally found on the stylus tablets. There has been one instance, so far, of the stylus form being found in the Vindolanda ink texts (ink tablet 601, line 3, see Bowman and Thomas (2003: 64). There is one other example of the expected ink form of E appearing as the stylus form in the pantheon of ORC texts—on an ink ostrakon from Mons Clandianus, text no. 304 (Bingen *et al.* 1997).

However, it cannot be ignored that the writing-surfaces of the two types of documents differ greatly (see S. 1.2). Earlier comparisons between writing on papyri with ink, and the inscribed writing of a stylus on wax indicate that there are differences in the process of writing between the two: 'In writing with a stilus upon wax there is a natural tendency to make short disjointed strokes' (Van Hoesen 1915: 3). Thompson talks of the 'smooth but clinging surface of wax scratched with the point of the stilus, or the less impeding... wood... inscribed in ink with the reed' (1912: 315), giving examples of the different letter forms found on documents at Pompeii:

we have writing on two kinds of material, and different accordingly; that of the deeds themselves, incised on the waxed pages with the stylus in decidedly cursive characters; and that of the endorsements and lists of witnesses, written in ink upon the bare wood of the pages which were not coated with wax, in a generally more restrained style and employing other forms of certain letters... the natural tendency, in writing on a resisting or clinging surface such as wax, is to turn the point of the writing implement inwards and hence to slope the letters to the left. The letters employed by preference, where a choice is possible, would usually be those which are more easily written in disconnected strokes, such as the two-stroke E and the four-stroke M... On the other hand, we find here the ordinary capital N... (Thompson 1912: 319)

3.1.2.1. Forensic palaeography and the stylus tablets

A few papers exist which investigate the effect different writing implements have on handwriting, a topic of interest in forensic science and criminology (Hilton 1959, 1984; Masson 1985). The most useful research is detailed by J. Mathyer (1969) who systematically compares ninety-eight standards prepared by fourteen persons using seven different writing implements. Mathyer aimed to investigate an hypothesis suggested by Edmond Solange-Pellat (1927): 'Le geste graphique est sous l'influence immédiate du cerveau. Sa forme n'est pas modifiée par l'organe scripteur si celui-ci fonctionne normalement et se trouve adapté à sa fonction.'⁶

⁶ Mathyer translates this as 'The graphic movement is under the immediate influence of the brain. Its form is not modified by the writing organ if this one works normally and is sufficiently adapted to its function' (p. 105).

Mathyer considers the role of the 'writing organ, ie the fingers and hand (right or left) for a person writing on a sheet of paper, the hand, forearm, the arm and the shoulder for a person writing at a black-board' (1969: 106), demonstrating that a change in the physiology of the writing process has little effect on the writing produced, indicating that it is a higher level cognitive process. He also graphically shows how a change in instrument does not affect a writer's handwriting, adapting the earlier hypothesis to his findings:

The form and line quality of the handwriting of a person is not modified by the writing instrument if this one works normally.

When the writing instrument does not work well... it can of course contract characteristics which indicate that the writer has tried with more or less success to obtain an acceptable result with a bad instrument... The influence of the instrument is rarely very important; this influence is frequently non-existent. (p. 106)

It can be presumed that, for the most part, the writers of the ink and stylus tablets *try* to use the same letter forms on each (as discussed in S. 2.1.3: scribes write letter forms which are intended to be read by others, and so they usually follow the conventions which differentiate one letter form from another). Mathyer's paper indicates that differences in letter forms between writing on wood with ink and incising the letters in wax should be minimal. Although there may be some variance caused by the different mediums, the scribes would have had the *intention* to write the same type of letter forms, and for the most part, these should be similar (aside from the letter E, as shown above—this, indeed, is demonstrated in S. 3.11). These small differences, however, would prove enough to throw off any existing image-processing techniques that could be used to scale, find, and match letter forms found in the ink tablets to those on the stylus tablets. Each individual usage of the characters is not identical in the ink tablets: it is human handwriting that is being dealt with here, not machine printed text. Instead, it was necessary to find a way of modelling the existing letter forms, which captured the types of knowledge the papyrologists and palaeographers employ when identifying and discussing individual letter forms. In doing so, this would provide a way of representing the letter forms contained within the

stylus tablets, and so aid in the identification of unknown characters which conform to or match those abstracted representations.

3.2. KNOWLEDGE ELICITATION AND PALAEOGRAPHY

The aim of this knowledge elicitation exercise was not to collect and summarize information regarding each individual letter form, which has been covered more than adequately in Bowman and Thomas (1983, 1994, 2003), but rather to enquire: how do the papyrologists identify individual letters? What are the characteristics and relationships of strokes that are noted by the palaeographers, which build up to make a character? Is there a way of modelling this information, so that a formal way of documenting characters can be developed, which would enable the information to be eventually processed by a computer (modelling being the first step in documenting data structures so that they can be processed automatically: De Carteret and Vidgen 1995). Developing an encoding scheme would also enable a training corpus to be constructed (the use of corpora in building trainable intelligent systems being a growing trend in the field of artificial intelligence, particularly in natural language processing (Charniak 1993; Lawler and Dry 1998), and computer vision (Robertson 1999; Robertson and Laddaga 2002)). A program of knowledge acquisition and elicitation (see S. 2.2) was undertaken to gain an understanding of the types of information the experts use when discussing and reasoning about letter forms.

3.2.1. Textual Sources

The major textual sources available which deal with the letter forms from Vindolanda are, of course, *Tab. Vindol. I* and *II* (Bowman and Thomas 1983, 1994). These were examined in detail, and formed the basis for the information collected about the letter forms. There were additional associated articles that proved helpful, by Cencetti (1950) and Casamassima and Staraz (1977), which deal with other sources of ORC, but detail the letter forms closely. Older sources, such as

Thompson (1912) and Van Hoesen (1915), although somewhat outdated, still provided useful information regarding the reading of letter forms.

A secondary source of information was generated from the text of *Tabulae Vindolandenses II*. All instances of discussions of individual letter forms from within the published commentaries of the Vindolanda ink texts (see S. 2.3) were collated, utilizing WordSmith (see S. 2.4.2). This provided instances of discussion regarding the identification of individual letters, particularly where this identification was in doubt. Likewise, every instance of discussion of individual letter forms from within the Think Aloud Protocols (see S. 2.3.2) was collated, providing information regarding the letter forms on the ink and the stylus texts.

3.2.2. First Stages in Knowledge Elicitation

To gain an understanding of the domain area, each expert was approached and asked to discuss at length the letter forms that are present within the Vindolanda texts. At first these were general discussions to gain an insight into the field. The experts were then interviewed using a semi-structured interview technique and asked to focus on specific issues, such as:

- individual characters
- standard forms of characters
- deviations from standard forms
- the physical relationship of characters to one another (ligatures, serifs, etc.)
- characters which are often confused with another
- specific features, such as ascenders and descenders
- the identification of similar hands in different documents.

This resulted in a large volume of data regarding the different characteristics of individual letters, and it was necessary to resolve this into a simpler form so an encoding scheme could be developed that would enable a corpus of letter forms contained in the Vindolanda texts to be constructed. This, in turn, could be used to train a system to aid in the reading of the stylus texts (as discussed in Chapters 4 and 5, and Appendix A).

3.2.3. Use of Repertory Grid

To aid in aggregating the data regarding characteristics of letter forms, a Repertory Grid was employed. Repertory Grids are based on personal construct theory (Kelly 1955), where individual concepts, or 'elements', are defined by the way they are alike or similar to other concepts. The identifying topics are the constructs in the relationship (see S. 2.2). By using a Repertory Grid to accumulate data regarding the Vindolanda letter forms, the main constructs become apparent, highlighting the most important characteristics of letter forms mentioned in all the available data. (It should be noted that the Repertory Grid was not used directly with the experts, which would be another possible means of capturing their knowledge regarding letter forms, as there were limitations on time and availability. In future, this may be another tool that can be used. Here, it was used by the knowledge engineer as a means of accumulating information from different varied resources.)

The program used was WebGrid,⁷ developed by Gaines and Shaw (1997) at the Knowledge Science Institute, University of Calgary (see Figure 3.3). WebGrid is a web-based knowledge acquisition and inference server that uses an extended Repertory Grid system for knowledge acquisition, inductive inference for knowledge modelling, and an integrated knowledge-based system shell for inference. In this case, it was used primarily for knowledge acquisition and modelling, to build up a detailed understanding of the constructs utilized when discussing the letters used in the documents at Vindolanda.

Once the data was captured in this manner, it was resolved into an overall schema which describes the types of information experts refer to when discussing and identifying individual letter forms.

3.3. INFORMATION USED WHEN DISCUSSING LETTER FORMS

The information gathered regarding the type of information used by the experts when discussing letter forms is shown in Figure 3.4. Each

⁷ <http://tiger.cpsc.ucalgary.ca/Webgrid/webgrid.html>

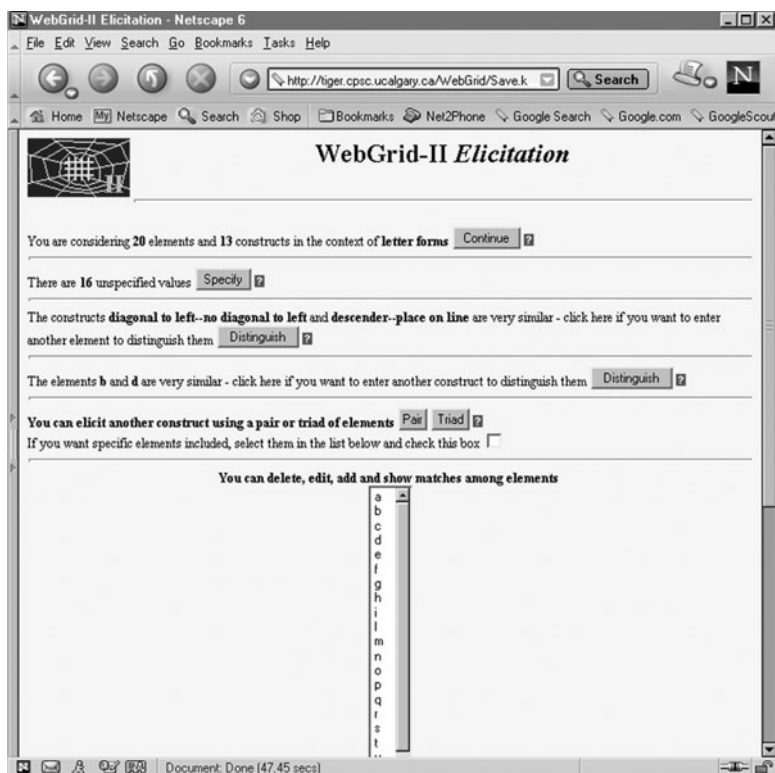


Figure 3.3. The WebGrid interface. The twenty letter forms were identified as separate ‘elements’ and the characteristics identified as ‘constructs’. The program identifies areas which need qualification; in this case it recognizes that the characters B and D are very similar, and asks the user to add another construct to distinguish them. In this way a list of constructs was built up which encompasses all the information included in the various sources.

character has one or more basic forms. These consist of one or more strokes, which combine to make a character. Character forms may differ depending on the context they are placed in, and may also be hard to distinguish from similar character forms.

There may be more than one basic form, or glyph, of a character (a glyph being the distinct visual representation of a character). One of these forms may be more usual or common than the others. There may also be a different capital form which is used in the texts. These may

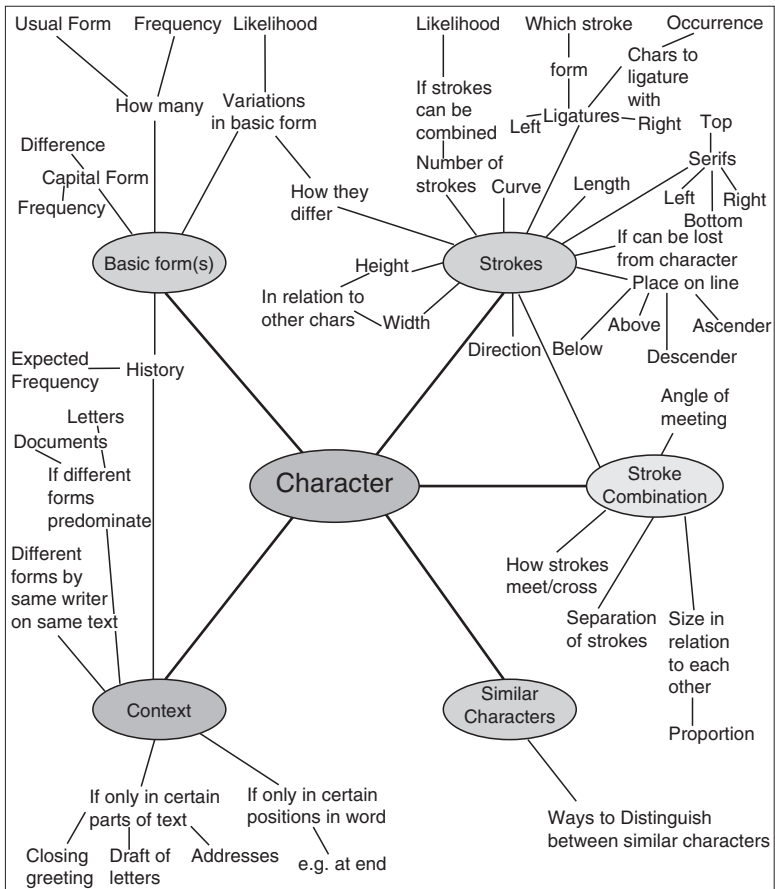


Figure 3.4. Information used when discussing letter forms, mapped as a semantic network, or ontology, to present a formal, explicit specification of the relationships between the information captured. A similar approach to the collection and representation of information in this manner which is then used to train an AI system can be found in Connell and Brady (1987).

differ substantially from the lower case letter, or may be very similar. The upper case character may be more or less frequent than the lower case. There may be variations on the basic form(s), which differ slightly on the stroke level. Some variations will be more or less common than others.

Each character consists of one or more strokes. Character forms will have a usual number of strokes, some of which may be combined

in the variations on the basic form. The strokes will be made in a certain order⁸ (it is difficult to definitely prove the order in which strokes were made, but assumptions can sometimes be made about the order of stroke data). It is occasionally possible to see where characters are made in a different way or order from the standard. Some strokes can be lost from the character in some variations in letter form: the fact that a stroke is missing may not hinder the identification of the letter in some cases.

Each stroke has a height, width, and length. These are relative, and can be compared with other strokes, possibly in other instances of the same character or even in other different characters within the same word, line, or document. Each stroke also has a relative direction and curve. Strokes (and characters) have a place on the base writing line: ascenders and descenders can give important clues as to character identification.

Strokes can also be (or have) ligatures and serifs. Ligatures join one character to another in a fluid writing motion. They may have a particular form from a certain stroke in a letter, and may be to the right, or from the left. Each character form has common letters that it will ligature with, and letters that it will rarely ligature with. A serif is a short decorative foot at the end of a stroke. They may be to the right or left, at the top or bottom of a stroke, and occur in some characters more often than others. Some character forms can be written with or without serifs or ligatures. Some characters never have ligatures or serifs. Stroke endings with no ligatures or serifs can either be blunt, or slightly hooked.

Strokes combine together to make character forms. Strokes which constitute a character will have a relationship to each other. How the strokes meet or cross is important: strokes may meet end to end, or cross over each other. Alternatively, there may be a gap, or separation, remaining between strokes. This meeting, or junction point (whether closed or open) will have an angle. The size of strokes, in relation to each other, can also be important, as the proportion of one stroke when compared to another can impart information which leads to a possible identification.

⁸ Mallon (1952) has a tendency to derive stroke order from character forms.

Each character form may vary depending on the context it is placed in. In some rare cases, it has been observed that letter forms vary depending on the position the characters have in the word: the final letter may take a different form than the same letter earlier in the word. Also, it has been suggested that where the letters are found in the text can have an effect on their form: there are types of letters which have only been found in closing greetings, drafts, or addresses. There may be a predomination of certain forms of letters in official documentation, which are rarely used in private letters (although this area requires much further research).⁹ The development of letter forms can be traced throughout different texts, and the context of the letter can give some indication as to the form which would most likely be expected, as some forms are historically used in certain circumstances. Different forms of letters are commonly found in the same text written by the same scribe, with no apparent reason for the difference in usage (a common feature of handwriting no matter which language is being used).

Some characters may often be confused with others. For example, the letters R and A are often confused because of the similarity of their form.¹⁰ When the use of specific features allowed these confused characters to be differentiated, they were noted.

Some additional information was collected regarding formatting which is not represented in Figure 3.4. For example, interpuncts are sometimes used as word separators. Indentation is also used in the texts to indicate where a section begins. Word spacing is also used (although not in all documents).

All of the above information has been discussed in relation to the Vindolanda letter forms, and can be used when trying to identify an unknown character. To be able to capture this data regarding individual letters, an encoding scheme was constructed, based on the above schema, which would allow images of the Vindolanda texts

⁹ It is certainly true of the so-called 'Chancery Hands' found in documents of the later Roman period, see Bischoff (1990: 63, with further references in n. 75).

¹⁰ It is this similarity that would throw a character-by-character pattern recognition system, as it would plump for what appears to be the most likely individual solution. Since the characters are so similar 'noise would dominate signal'—i.e. small changes would sway the interpretation. In the approach adopted in this monograph, the disambiguation is left to higher level knowledge, increasing the accuracy of the interpretation, as discussed and demonstrated in Chs. 4 and 5, and Appendix A.

to be annotated in a manner that would encapsulate important character information.

3.4. DERIVED ENCODING SCHEME

Before being able to annotate any images, it was necessary to restructure the complex relationship of information discussed in regard to character forms, to provide a more linear structure, or markup scheme, consisting of broad headings that could easily be applied to images and their constituent parts. The markup scheme consists of over thirty different elements which can be used when annotating individual character strokes and can be summarized as follows:

Each area of space in the image is either a

- Character box (the area surrounding a collection of strokes which make up a individual character)
- Space character (indicating the empty space between words)
- Paragraph character (indicating areas of indentation in the text)
- Interpunct (indicating the mark sometimes used to differentiate words in this period of ORC).

Each letter is comprised of strokes

- Traced and numbered individually—this numbering being used to identify separate strokes, and not necessarily reflecting the order in which the strokes were made, as this is very difficult to determine in such abraded texts. Identifying strokes in this manner involves some interpretation by the annotator regarding what is actually present, and what can be identified as an individual stroke.

Each stroke has end points, either

- blunt (i.e. with no complex formatting to the end point)
- hooked (with direction specified, i.e. up left, up right, etc. The direction specified was relative to the base line of the text in the document.)
- ligature (with direction specified)
- serif (with direction specified).

Each character may also have stroke meetings, where strokes meet, combine, or join. These are either:

- end to end (where the ends of strokes meet, such as in our letter ‘N’)
 - exact meet (where the ends of strokes meet exactly)
 - close meet (where there is a small gap between the ends of the two strokes meeting)
 - or crossing (where there is a small overlap between the ends of the two strokes meeting),
- or middle to end (where one stroke meets the middle of another, such as in the junction in our capital letter ‘T’)
 - exact
 - close
 - crossing,
- or crossing (middle to middle, such as with the letter ‘X’).

These were the most important set of characteristics, and were incorporated into image annotation software (see below 3.6) to be able to encode this data graphically. However, there was much more information available regarding letter forms. For example, one of the most important characteristics which helps identify letters is whether they have strokes ascending above or descending below the writing line; the letter S is often easily recognizable because of its long descenders. A further collation of all features and types of strokes mentioned by the papyrologists as they discussed the texts was undertaken, and resolved into a textual schema. This was used to provide additional tags to the annotated regions to help in labelling them further. These tags were added manually, in tandem with the annotations made with the graphical interface.

Each character box had a letter identification (*) and an overall size height (SH) and width (SW). Letters that had not been confidently identified by the papyrologists were marked with a question mark (*?). Letters which were expected by the papyrologists but missing due to damage of the texts were marked with a character box (with no strokes included) but annotated with an exclamation mark (*!). For example ‘*s?’ indicates that the character box contains all the strokes regarding a letter which may be the letter S, although this was not a confident identification.

Each stroke was assigned additional textual tags, having:

- A direction (Direction), giving the stroke orientation and type, which included

- Straight strokes (DirectionStraight).
- Simple curved strokes (DirectionCurved).
- Complex curved strokes (DirectionCurvedWave).
- Loops (DirectionLoop): each of these tags included further orientation tags to indicate left, right, etc (orientation being dictated by taking the centre point of the character box on the base line of writing, and deriving from that up, down, left, right, etc. Describing orientation in this manner replicates the language the papyrologists use when discussing texts, and also includes some level of abstraction in the modelling of the characters, which is useful for comparing descriptions of the characters later. If the direction had been measured absolutely, for example using co-ordinates, it would be very difficult to compare individual instances of the characters automatically, as the information captured would have been too specific, a problem known in engineering science as ‘over-fitting’ of data.) For example, DirectionStraightDownLeft represents a stroke which is straight, and veers to the left, down from the centre of the character box, away from the centre.
- A length (Length), being either comparatively short, average, or long. (Again, this replicated the language used by the papyrologists when discussing stroke length).
- A width (Width) being either comparatively thin, average, or wide. (Replicating the variables used in the papyrologists discussion of stroke width).
- A place (Place) on the writing line, being either within the average, descending below the nominal writing line, or ascending above the average height of a character.

Each stroke meeting, or junction, had

- an angle (A), with note taken of the orientation of the meeting and whether the angle was acute, right, obtuse, or parallel.

If the strokes were broken (due to damage of the document) this was noted in the stroke ending field, where they were labelled as being blunt with the additional ‘broken’ textual tag added.

A full textual representation of this encoding scheme is available in Appendix B.

3.5. BUILDING THE DATA SET

Seven ink tablets were identified,¹¹ with the help of one of the experts, which would provide good, clear images of text to annotate, and have enough textual content to provide a suitable set of test data. Of these seven, 225b and 225f have been identified as being by the same scribe (being two sides of the same document), and 248 and

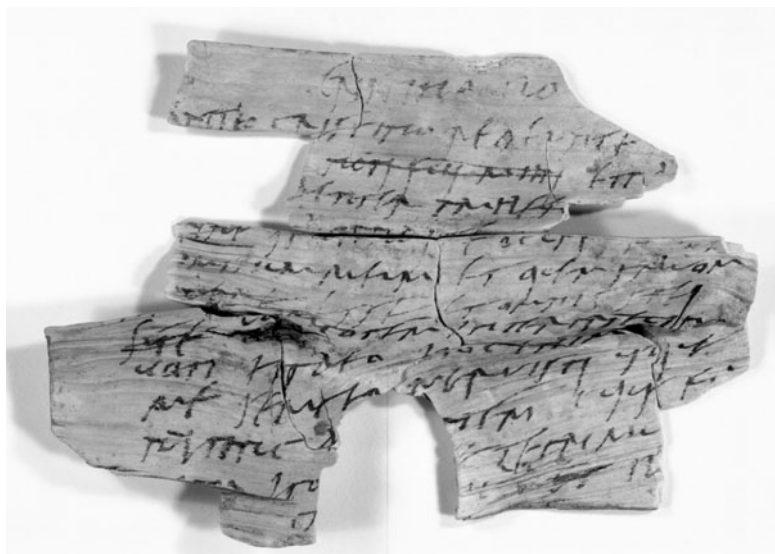


Figure 3.5. Ink text 225f. The front of a document. Although displaying considerable damage, this document displays a selection of type and form of characters. The style of handwriting is confident and inelegant, letters are joined boldly and more frequently than other hands, many of the ordinary cursive forms appear in an idiosyncratic way, and spaces are often left between words. The document is a draft of letter from Cerialis (prefect of the Ninth Cohort of Batavians at Vindolanda during period 3—c. AD 97–102/3) to Crispinus, who may have been a higher placed official. Cerialis is writing to ask Crispinus to use his influence to make Cerialis' Military service pleasant by putting him on good terms with as many influential people as possible. A full discussion of this text can be found in Bowman and Thomas (1994: 200; Bowman 1983: 41).

¹¹ 225b, 225f, 248, 255, 291, 309, 311 from Bowman and Thomas (1994).

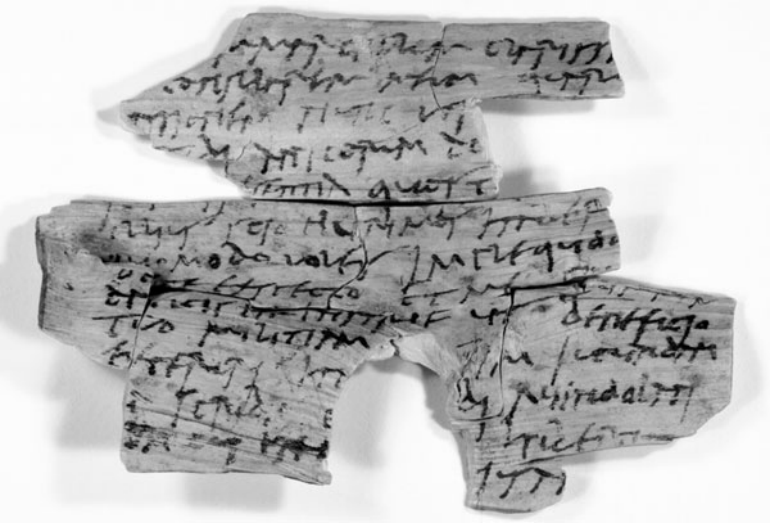


Figure 3.6. Ink text 225b, the back of the letter shown in Figure 3.5. This tablet contains variation in letter forms and is a good source for the training set.



Figure 3.7. Ink text 248. A more clearly written letter on a complete leaf in two columns, with different character forms from the tablets above. Note the long ascenders and descenders. The scribe uses a deliberate apex over some instances of the letter o. This is a letter from Niger and Brochus to Cerialis, giving good wishes, and wishing good luck and success regarding his planned actions (see Bowman and Thomas 1994: 219, for full transcription).

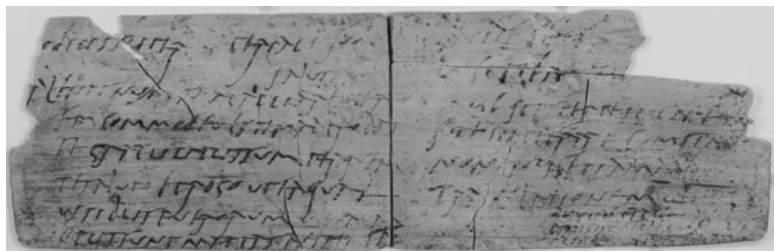


Figure 3.8. Ink text 255. A complete leaf in two columns, containing letter forms in a good, regular cursive hand. This hand uses two forms of the letter d, and the letter l, a very idiosyncratic u (which is almost a flat dash), and a form of the letter c which is like the letter p. This is a letter from Clodius Super, who may be a centurion, to Cerialis: the writer was pleased that Valentius, who had just returned from Gaul, had approved some clothing, and requested from Cerialis some cloaks and tunics for his boys (either slaves or sons). See Bowman and Thomas (1994: 225) for full discussion.

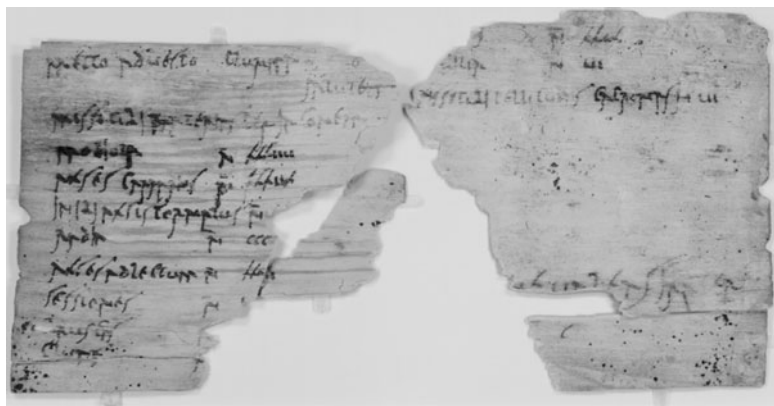


Figure 3.9. Ink text 309. A number of fragments make up this text, although the text is mostly complete. The hand is basic and rather inelegant. It is a fragmentary inventory of components of carts, planks, and boards. See Bowman and Thomas (2004: 286) for a transcription.

291 were also identified as by being by the same hand (although this is different from the scribe who wrote 225). The three other ink texts are in different hands. Additionally, closing valedictory comments are present in some of the documents written by the hand of the author rather than the scribe, providing further different instances of

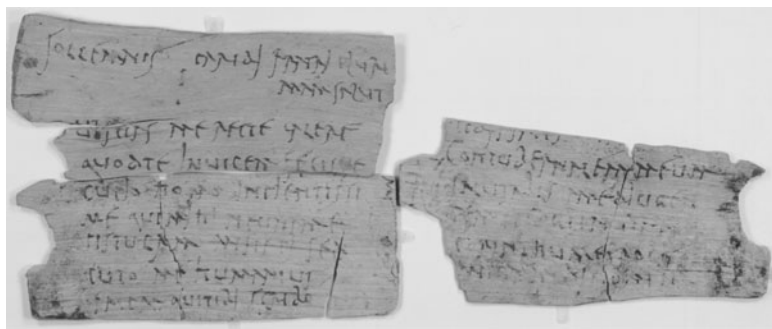


Figure 3.10. Ink text 311. Three joining fragments of a diptych. The main hand is a very competent and interesting squarish cursive, with carefully spaced words, and little use of ligatures. The writer also uses unusual forms of the letters e, u, and d. This is a letter from Sollemnis to Paris, complaining that Paris has not written to him, and sending greetings to three friends. See Bowman and Thomas (1994: 294) for full transcription.

letter forms. This sampling provides enough data to ensure that different letters forms and styles, derived from different authors, are covered in the training set.

The final text to be annotated was ink text 291, illustrated in Figure 1.2. The diptych is written in a slim, elegant script. The letters have marked ascenders and descenders and very little use of ligature. There is occasional use of the apex mark.

These seven texts are mostly personal correspondence (225b, 225f, 248, and 255 being taken from the correspondence of Flavius Cerialis, the prefect of the Ninth Cohort of Batavians, and 291 from the correspondence of his wife, Lepidina. 311 is a letter from Sollemnis to Paris, who was possibly a slave of the commander). The only example that is not a letter is 309, being the account of supplies (mostly manufactured wooden items) which were being delivered to Vindolanda. Although it is expected that the stylus tablets will contain slightly more formal subject matter, the difference in subject matter between the ink and stylus tablets should not matter for this corpus as the reason for constructing this training set was to provide a test set to compare letter forms, not words.

Two stylus tablets were also annotated, 974 and 797. The latter is a fragment of a personal letter (see Figure 2.3), 974 possibly concerns

some transactions involving the manumission of a slave (see Figure 2.4). These were chosen simply because they are two tablets that the experts have managed to read (although there are over 200 stylus tablets from Vindolanda only a few of them have been read so far, see S. 1.2). The last few lines of 974 have been read, providing examples of letter forms contained on the stylus tablets. The left and right hand columns of 797 elicited a few words, giving a small sample of characters. This gives a set of test data to see how effectively data regarding the character forms from the ink tablets can be used to aid in the reading of character forms from the stylus tablets (see Chapter 5 for test data and results).

One expert was provided with large images of the chosen texts, on which he traced the individual strokes of each letter as they appear on the ink tablets. These images were married up with the published texts (to confirm the text that had actually been read; letters such as R and A, or C and P, are easily confused) and the information gleaned from the knowledge elicitation exercises regarding character forms and formation to provide the information with which to annotate the images. Although this does involve some interpretation of the data by the knowledge engineer, it was done in as methodological and systematic a fashion as possible. The expert's tracings of the character forms were taken as the main source of information regarding the texts. Care was taken to annotate the images without the addition of personal bias by retaining a distanced stance of annotating exactly what was there, rather than what *should* be expected. The annotations were also double checked by cross referencing, as discussed below (S. 3.7.2).

3.6. THE USE OF THE GRAVA ANNOTATOR

To be able to annotate the images using this encoding scheme, an annotation program was used. This was an adjusted version of Paul Robertson's GRAVA Annotator Program, a Motif¹² program

¹² Motif is the industry standard graphical user interface environment for standardizing application presentation on a wide range of platforms. Developed by the Open Software Foundation (OSF, see www.opengroup.org) it is the leading user interface for the Unix system. In this case, we utilized Exceed, an X-Server for the PC, to run our Motif application on the Windows platform. More information about Motif can be found at <http://www.motifzone.com/>

originally developed to manually segment and label a corpus of aerial satellite images¹³ (see Robertson, 2001; Appendix C). The results of the annotation are written to a file in Extensible Markup Language¹⁴ (XML) format, to allow easy interaction with other programs. Robertson's software was adapted to incorporate the encoding scheme, detailed above, that was relevant to the Vindolanda texts, to allow these to be labelled in the same manner. This is the first known instance of textual encoding in the humanities on a stroke-based level: textual markup in the arts usually revolves around the marking up of text from the character level upwards, rather than marking up constituent strokes of characters themselves (Terras and Robertson 2004), and this serves as an example of how the adoption of techniques from artificial intelligence and computer and engineering science can further the remit of computing in the humanities.



Figure 3.11. The GRAVA annotation program. The facility allows regions to be hand traced and assigned a label. The application screen is divided into a menu bar, image display pane, command input area, and message area. Most interaction with the annotator is performed using the mouse.

¹³ The corpus produced in this case was subsequently used to train a system to segment, label, and parse aerial images so as to produce an image description similar to that produced by a human expert (Robertson 2001).

¹⁴ XML is a cross-platform, software- and hardware-independent, text-based markup language designed to describe data, so that it can be structured, stored, and manipulated. The markup tags must be defined, using a Document Type Definition (DTD) or XML Schema, hence the name Extensible, as XML tags are not predefined. Originally designed to meet the challenges of large-scale electronic publishing, XML is playing an increasingly important role in the exchange of a wide variety of data on the Internet. <http://www.w3.org/XML/>

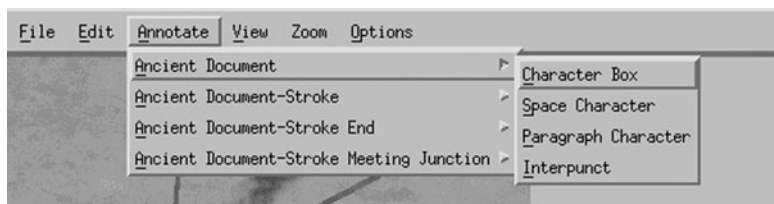


Figure 3.12. Drop down menus in the GRAVA Annotator, incorporating the encoding scheme determined from the knowledge elicitation exercises. This tool can be adapted to incorporate any structured encoding scheme, in order to label images with assigned information.

3.7. ANNOTATING THE CHARACTERS

Characters were annotated by first drawing around the outline of a character with the mouse, and assigning it a character label. Individual strokes were then traced, and numbered by selecting the option from the drop down menu. Stroke ends were then identified and labelled. Finally, stroke junctions were noted and assigned labels. An example of this is shown in Figure 3.13, using the letter S from the start of 311 as an example.

These annotations are preserved in an XML file which textually describes the annotated image. For example, this is the XML which describes the bounding character box of the letter 'S' in the below example:

```
<GTRegion author="Melissa Terras" regionType="ADDCHARACTER"
regionUID="RGNO" regiondate="04/04/02 18:07:16" coordinates="112,
142, 106, 223, 87, 317, 52, 377, 56, 420, 126, 423, 180, 349, 203, 244, 199,
108, 269, 71, 323, 28, 298, 7, 190, 23, 118, 69, 112, 142"></GTRegion>
```

The region type tag "ADDCHARACTER" defines this region as a character box which surrounds a collection of strokes, the region ID "RGNO" gives the annotated region a unique identification number, and the co-ordinates preserve the shape of the character box. Each individual region that is annotated (characters, strokes, stroke endings, and stroke meetings) has its own similar line in the XML file, with each having a unique identifying region ID, and a region type which specifies what type of annotation has been made. This file

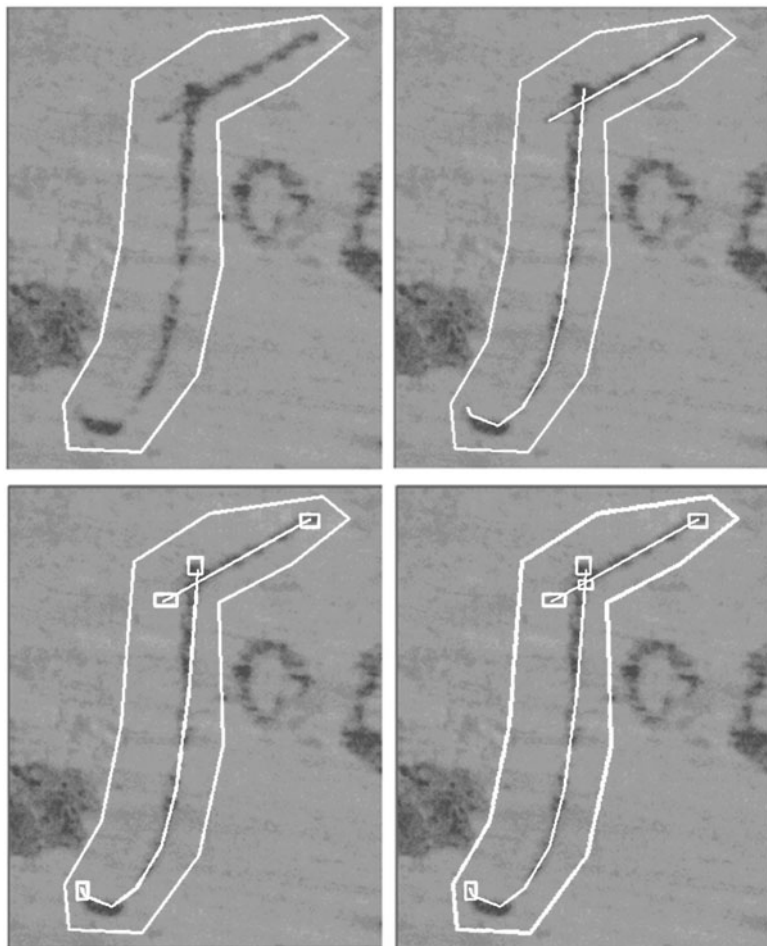


Figure 3.13. Steps taken in annotating a letter. The outline is traced at first, followed by individual strokes, stroke endings, and stroke meeting junctions. All are assigned labels from the drop down menus in the program.

is structured hierarchically; all strokes, stroke endings, and stroke meetings ‘belong’ to an individual character.

The structure of the XML generated is fully described in Appendix B.2. A table of region identifier codes is available in Appendix B.3.

3.7.1. Additional Annotations

The additional tags described in 3.4 were included into the annotation by assigning a textual code for each region in the ‘comments’ field of the GRAVA annotator. The letter S from 311, above, was deemed to be of large height and width, and so the additional comments added to the XML regarding the character box were:

comments=“*S, SHL, SWL”

“*S” identifies the character as the letter S, SHL stands for SizeHeightLarge indicating that the height is large, and SWL represents SizeWidthLarge, indicating that the width is large. These short comment annotations were chosen to increase the speed in annotation. This gives the final resulting XML output for this region as

```
<GTRegion author="Melissa Terras" regionType="ADDCHARACTER"
regionUID="RGN0" regiondate="04/04/02 18:07:16" coordinates="112,
142, 106, 223, 87, 317, 52, 377, 56, 420, 126, 423, 180, 349, 203, 244, 199,
108, 269, 71, 323, 28, 298, 7, 190, 23, 118, 69, 112, 142" comments="*S,
SHL, SWL"></GTRegion>
```

Individual strokes are then annotated in the same way, the XML data nesting in the ‘character box’ the strokes are associated with, and additional annotations can be added regarding individual strokes and junctions by inserting textual descriptions. All comment tags are provided in Appendix B.1. An example of a full XML file which describes this letter S is presented in Appendix B.2, including the annotations of individual strokes which constitute the letter. (It may be asked why TEI¹⁵ compliant XML was not used. First, the tagset for TEI does not include stroke information, although it could be expanded to include the elements identified here. The short comment tags were chosen for speed in annotation: transformations could be used on the XML files to retag the data in a more compliant manner, using each individual element as a tag, rather than listing them in the comments field, as they are captured in the existing data structure. This will make the resulting XML more human readable.)

¹⁵ The application of XML (Extensible Markup Language) developed by the Text Encoding Initiative. See <http://www.tei-c.org/> for more information.

3.7.2. Practicalities

Due to the fact that the images were very large, each document was split into a series of smaller images to allow easier annotation, on a line by line basis. Splitting the images in this way was also necessary due to the processing power needed to run the annotation program. There were 89 lines of ink text, and 21 lines of stylus text to annotate, resulting in 110 images in total.

At a later date, the files of annotations were cross-referenced with the texts printed in the published volumes to ensure that the annotations were correct. Any mistakes or areas of confusion were corrected, and the corpus updated. For example, the text 'USSIBUS' had been incorrectly annotated as 'USSIBUSS' in a section of 255. This was corrected. In this case, a version with the incorrect annotation was retained to allow testing on the system, as described in section 5.2. It is acknowledged that the annotations could contain some further examples of human error, but the annotations were carried out as systematically as possible, and this cross-referencing ensured that the majority of mistakes would be identified.

3.8. RESULTS

Each image was annotated in the way described above. Care was taken to be as methodological as possible whilst undertaking this task. In total, there were 1,506 individual characters from the ink tablets annotated, and 180 characters from the stylus tablets. The nine completed sets of annotated images represented approximately 300 hours of work, with an average of around six or seven characters being completed in an hour. The images were annotated and checked over a period of three months. The resulting corpus can be viewed using a Java-enabled web browser, as described in Appendix B. It will also eventually be displayed as part of the Vindolanda tablets online website,¹⁶ alongside readings and interpretations of the texts, which is hosted by the Centre for the Study of Ancient Documents, University of Oxford.

¹⁶ <http://vindolanda.csad.ox.ac.uk>



Figure 3.14. Completed line of annotated text, from ink tablet 255, annotated with ‘*ussibus puerorum meorum*’, meaning ‘for the use of my boys’, referring to some cloaks and tunics for Clodius Super’s *pueri* (Bowman and Thomas 1994: 225).

Each individual character from the corpus is also represented in Appendix C, giving a full palaeographic representation of all of the character forms found within these nine texts.

3.9. GATHERING CORPUS TEXTUAL DATA

The Vindolanda ink tablet corpus is the major contemporaneous resource to the Vindolanda stylus tablets, and as well as being used to generate a corpus of letter forms, can be used to generate other textual data regarding Latin of this period. This information is necessary to train the system described in Chapters 4 and 5, and Appendix A, and so further linguistic information was extrapolated from the Vindolanda ink tablets corpus.

The word list generated from this corpus may be different from that produced from any other available source material regarding Latin, due to the temporally and geographically distinct nature of the corpus.¹⁷ Although the language used in the texts is not strictly Vulgar

¹⁷ ‘We can even talk of regional “dialects” of Latin. We can suppose this partly on the basis of generalities that have been discovered to be empirically true for the study of all languages; when a language is used of a wide and disparate geographical area, influenced by widely varying external factors of an ethnic and socio-cultural type, geographical variants can arise that are noticeably different from each other despite the fact that they all form part of a single linguistic system . . . It seems likely that in imperial times a slight amount of geographical variation did slowly arise in Latin, affecting pronunciation in particular, but perhaps also a few morphological details (ignoring, of course, the wide differences we find in personal and place names, due to the different ethnic origins of the populations of different areas.) This kind of divergence posed no threat to the

Latin,¹⁸ 'they contain . . . a stock of terms . . . of the type which are rarely, if at all, attested in literary genres' (Adams 1995: 120). The texts contain lexical and syntactic 'errors',¹⁹ and utilize new words, new meanings of known words, the first attestations of abnormal forms of words, Celtic loan words, and anticipations of Romance language. Most other large sources of Latin words and grammar deal with Classical Latin: a version of the language that was written by a few authors, and spoken by almost no one; Vulgar Latin was spoken by millions, and the Vindolanda texts are one of the only large sources available on which to base any conclusions regarding the language. For these reasons, the entire known corpus of ink texts from Vindolanda (as of December 2001) was the only source used to generate lexicostatistics to aid in the reading of the stylus texts. In the future, it will be possible to generate other sets of statistics from different corpora, which could be used to analyse how similar the Vindolanda corpus is to other texts available (as these documents are correspondence from the Roman army, there must be standardization in the language used, so that individuals from different outposts could communicate—statistical comparisons made between letter frequency from the Vindolanda text and the Perseus corpus (S. 3.10) would seem to indicate that this was the case). This additional data could also be used to provide alternative information sources for the system described in this monograph. However, the Vindolanda corpus was the only source used here as it was readily available, is contemporaneous to the texts found on the stylus tablets, and is of a large enough size to provide statistics with which to test the implementation of the system.

The Vindolanda ink corpus used comprised of the 230 Latin texts published in Bowman and Thomas (1994), plus 56 new texts that had been read in preparation for the publication of the next volume of the Vindolanda texts (Bowman and Thomas 2003). There were 27,364 characters (excluding space characters) in total, comprising

fundamental unity of the language, which is hardly surprising in view of the centralising power of the empire itself and the strength of its traditions' (Herman 2000: 116–19).

¹⁸ 'By *Vulgar Latin* is meant primarily that form of the language which was used by the illiterate majority of the Latin-speaking population' (Coleman 1993: 2). For a discussion of the relation of the language of the Vindolanda texts to Vulgar Latin, see Adams (1995: 131).

¹⁹ When compared with literary Latin.

6,532 words, or word fragments. In the corpus there were 2,433 unique word tokens (1,801 words appeared only once, the rest were repeated). This should provide an adequate corpus on which to base any conclusions about the language used at Vindolanda.²⁰

This corpus was analysed using WordSmith (see S. 2.4.2) and TACT,²¹ two common corpus linguistics computer programs, to generate the following data:

- frequencies of letters
- letter combinations (bi-graph analysis)
- word lists
- frequencies of words.

These statistics were used to train the artificial intelligence system described in Chapters 4 and 5, and Appendix A.

3.10. HOW REPRESENTATIVE IS THE CORPUS?

Due to the fragmentary nature of the Vindolanda corpus, it is difficult to be sure how representative the coverage is, of the words and letters in the text corpus, and the characters within the annotated image corpus. There are relatively few texts that the Vindolanda corpus can be compared with on a word level, as textual evidence from this period of the Roman Empire is rare. It is also difficult to obtain the necessary statistics to allow any comparisons. However, on a broader level, it is possible to compare the Vindolanda corpus with the largest corpus of written Latin, that of the Perseus Digital Library.²²

²⁰ The representativeness of a corpus has been much discussed in the field of corpus linguistics (Kenny 1982; Biber 1983, 1990; Oostdijk 1988), the consensus being that 'small is beautiful' and that 'there is every reason to make maximal use of these corpora [of 2,000-word length] for analysis of linguistic variation until larger corpora become readily available' (Biber 1990: 269). The major corpus of American English (known as the Brown Corpus, see Francis and Kuchera 1979) uses 2,000-word samples to represent its various genres. Zipf's 'law' (1935), which predicts an 'ideal' set of frequencies (and hence probabilities) for lexical items given a particular vocabulary size, maintains that a dictionary size of 1,000 words will give a percentage cover of 56.220% of the language, although this is based on an analysis of English-language texts. In the case of this research, the corpus used is 100% of the known corpus of Latin for this period.

²¹ 'Text Analysis Computing Tools': <http://www.chass.utoronto.ca/cch/tact.html>

²² www.perseus.tufts.edu

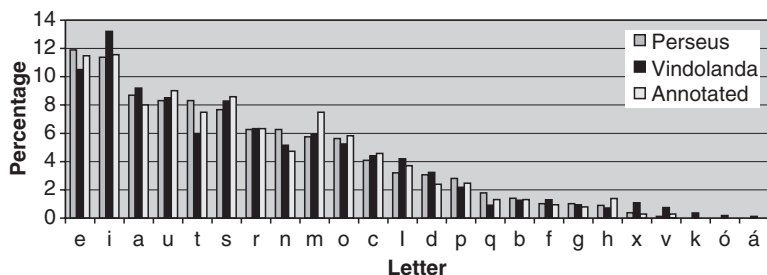


Figure 3.15. Comparisons of letter frequency in the Perseus Digital Library, the Vindolanda text corpus, and the annotated set of images of the Vindolanda ink and stylus texts.

The Perseus Digital Library maintains a textual corpus of almost 7.8 million characters which provides comprehensive coverage of Classical Latin. The corpus is primarily comprised of classical commentaries and histories.²³ Although the ‘Vulgar’ Latin in the Vindolanda corpus differs semantically and grammatically from Classical Latin, on a letter by letter basis the frequencies of characters present should be very similar (as indicated by Zipf (1935) in his comparison of letter frequencies in the English Language). A comparison between the letter distribution in the Perseus data set and the Vindolanda corpus can then be taken as a rough indication that the coverage of the image and the textual corpora are adequate. Statistics regarding the letter frequencies in the Perseus corpus can be found in Mahoney and Rydberg-Cox (2001). These were compared to the letter distribution in the Vindolanda text corpus, and to the distribution of letters in the annotated image corpus.

Even though the Perseus corpus has 7.8 million characters, the Vindolanda text corpus (see 3.9) 27,000 characters, and the annotated image corpus only 1,700 characters, the frequencies are remarkably similar. This can be taken as an indication that the annotated characters give a good distribution of the letter forms used in the documents, and that the lexicostatistics derived from the Vindolanda corpus will have some usefulness and relevance to our task.

²³ The database consists of the following texts: Plautus, Caesar (*BG*), Catullus, Cicero (orations and letters), Virgil, Horace (*Odes*), Livy (books 1–10), Ovid (*Metamorphoses*), Suetonius (*Caesars*), the Vulgate, and Servius’s commentary on Virgil.

3.11. LETTER FORMS

It was suggested in section 3.1.2 that the letter forms contained within the Vindolanda stylus texts should be similar to those contained within the ink documents (this being the assumption that the palaeographers work from). The data set constructed during this part of the research contains enough information to evaluate whether this is the case. Appendix C contains all the representations of the characters used within this research, and every instance of the characters in the ink and stylus tablets. Although the author of this monograph is not a palaeographer, there are some observations that can be made regarding the variations in letter forms between the ink and stylus texts from the data given in this Appendix.

- A. On the whole, the character A contained within the ink and stylus texts seems to be very similar. In some cases the two strokes meet at a sharper angle in the stylus A, with the first stroke being more upright than in the ink version.
- B. The ink B is fairly fluidly made, whilst those found in the stylus texts are less curved. The second, longer stroke in the stylus B tends to be more straight and does not have the pronounced loops to each end of the stroke as present in the ink B.
- C. The letter C is very similar in both texts. In some cases of the stylus text the loop can be less full than in the ink text, with the length of stroke also being shorter in the stylus C.
- D. The longer stroke of the stylus D tends to slope top left to bottom right, whereas that of the ink D can be more variable. As with the letter B, there tends to be less curvature in the long stroke of the stylus D than the ink D.
- E. As already noted (S. 3.2), the ink and stylus texts contain a different form of the letter E. This can be seen clearly in the examples given. There are no examples of the stylus form being used in the ink texts, or the ink form being used within the stylus texts, in the data set that was constructed. That is not to say that this would never be the case, however.
- F. There were no examples of the letter F in the stylus image corpus.

- G. As there was only one example of a letter G in the stylus image corpus, it is impossible to make any generalizations regarding the differences between the ink and stylus letter G. However, this one instance is very similar to those found in the ink texts.
- H. There were no examples of the letter H in the stylus image corpus.
- I. The stylus I shows less use of ligatures and serifs than that found in the ink texts. The models generated from the ink and stylus texts show that the ink I tends to slope from bottom left to upper right, whilst the stylus I slopes from upper left to bottom right. This is pronounced in the models derived from the corpus (see SS. 4.4 and A.6) due to the fact that the strokes in each case were almost vertical, and they have been stretched to fit into the canonical 21 by 21 pixel grid. Nevertheless, this does indicate that in general the stylus I tends to slope a little more in the opposite direction to the ink I.
- L. The ink L shows a lot of variation in direction, angle, and use of serifs. The small number of examples of stylus L available indicate that this variation is also present in the stylus tablets. However, there seems to be less use of serifs in the stylus tablet forms.
- M. The ink M shows considerable variation in its form, width, and angle of stroke meetings. This variation is echoed in the stylus tablet form. The stylus M seems to be less fluid, and the individual strokes more separated than those in the ink M, which can often run together smoothly.
- N. The ink N shows considerable variation in its size and angle of stroke meetings. This is reflected in the forms of the stylus tablet N.
- O. The ink text O is commonly made in one stroke, which makes a loop. There are a couple of instances where it is made with two strokes in the corpus. The stylus tablet O is made with two combining strokes in every example in the corpus, and never made in one single stroke.
- P. The ink tablet P shows some degree of variation. There is only one example of a stylus letter P, and so comparisons between the two are limited. However, the stylus letter P presented does seem to be very similar to the ink letter P.
- Q. The longer stroke of the ink letter Q tends to be more upright than that of the stylus Q, which tends to slope more from upper left to lower right. The bow of the ink Q tends to be more of a more elongated width and smaller height than that of the stylus Q.

- R. The ink R shows some variation, but the first stroke seems to always be the longer, with the second stroke sitting above the first. In the stylus tablets, there is considerable variation in the letter form. Sometimes both strokes are of the same length, and can meet end to end, rather than overlapping. The first stroke tends to be straighter in the stylus tablets than that of the ink tablets.
- S. The ink S shows some variation in form, but in general the form found in the stylus tablets echoes that found in the ink texts.
- T. Again, the form found in the ink texts shows some variation, and this is the case with those found in the stylus tablets, but in general the forms found on both texts are very similar.
- U. U and V are somewhat interchangeable in the texts. In many cases of the ink text, the character is made in one fluid motion, with one curved stroke. However, in the stylus text, the U is always made with two strokes, and these can often be straight, rather than curved, indicating the difficulty on curving the stylus through the wax surface.
- V. These letter forms appear to be similar, but as with the U, above, the second stroke on the stylus text V is often very straight, and there is no instance of the character made in one fluid motion with only one stroke.
- X. There were no examples from the stylus tablets with which to compare the forms from the ink texts.

On the whole, the letter forms used on the stylus text do seem to be similar to those contained within the ink documents. The main difference between the two character sets involve strokes which are looped or curved: it is much more difficult to curve the stylus through the wax than make the same fluid motion with ink on wood. However, the forms of the character remain the same (apart from the letter E, as demonstrated), and so models of the ink characters, and the stylus characters available, should provide an adequate means of reading the unknown characters contained within the stylus texts in the future. How this is achieved is explored in Chapter 4, with more technical details provided in Appendix A, and results of the system demonstrated in Chapter 5.

3.12. CONCLUSION

In this chapter, a review of palaeographic research regarding Old Roman Cursive was presented, and a series of knowledge elicitation exercises carried out, to enable the encapsulation of the types of information experts utilize when discussing and identifying letter forms. This enabled an encoding scheme to be developed so that a corpus of annotated images could be constructed, annotating images of text to produce XML representations of letter forms. The resulting data set is the only such amalgamation of palaeographical information regarding Old Roman Cursive handwriting in existence, and as such, is a unique resource of this period for papyrologists and palaeographers. This corpus has already provided the means to investigate the difference between the letter forms as found on the ink texts and the letter forms found on the stylus tablets. The corpus of annotated images, combined with the linguistic corpus generated from the Vindolanda ink texts, was utilized as the means to train a system to attempt to read the ink and stylus tablets, as discussed in Chapters 4 and 5, and Appendix A.

The knowledge elicitation techniques used in this chapter were time-consuming, and the amalgamation of knowledge depended on the thoroughness of the knowledge engineer. However, the resulting encoding scheme is comprehensive. Although it refers to Old Roman Cursive in particular, it could be used as the basis of a scheme to annotate any stroke-based written text. There were some elements of the encoding scheme that were felt to be not as relevant as others when it came to annotating the Vindolanda texts (for example, stroke width is fairly standardized throughout the corpus) but the encoding scheme provides the means to encapsulate different types of graphical information. Often, the relevance of the information only becomes apparent after the annotation has taken place, and this encoding scheme and annotation tool enables as much information as possible to be captured in the image corpus.

Annotating the corpus was a tedious, time-consuming task. Although it was carried out in the most systematic way possible, there may very well be some undetected human error included in the annotations. One way to resolve this problem would be to annotate

the files again, and to compare the resulting annotations: any areas of difference would be highlighted, and the confusion resolved. Fortunately for the author, there was not enough time available during the research to enable this to happen.

It has been demonstrated here that it is possible to develop and implement tools which allow the capturing of complex palaeographic data through marking up images of text to generate XML representations of letter forms. Continuing the development of such schemes and tools will increase the usefulness of computing tools for the palaeographic community, and will provide the means to train systems to read previously illegible ancient texts. Adapting techniques from artificial intelligence has thus made it possible to extend the remit of 'textual markup' in the humanities, marking up constituent strokes of characters themselves rather than merely marking up text from the character level-upwards.

This page intentionally left blank

II

Image to Interpretation

This page intentionally left blank

Using Artificial Intelligence to Read the Vindolanda Texts

Co-authored with Dr Paul Robertson

The story was that [Theuth] was the first to discover numbers and calculation, and geometry and astronomy, and . . . to cap it all, letters . . . The story goes that Thamus expressed many views to Theuth . . . when it came to the subject of letters, Theuth said ‘But *this* study, King Thamus, will make the Egyptians wise and improve their memory; what I have discovered is an elixir of memory and wisdom.’

Thamus replied ‘Most scientific Theuth . . . you have discovered an elixir not of memory but of reminding. To your students you give an appearance of wisdom, not in the reality of it; having heard much, in the absence of teaching, they will appear to know much when for the most part they know nothing.’

(Plato, *Phaedrus* 274–5 (1986: 123))

The research presented so far in this monograph has two usable (and useful) products. First, the proposed model of how papyrologists operate (S. 2.8.3) can be used as the basis for the design of a computer system that can be trained to read the Vindolanda texts: the process has been broken into individual modules, or ‘agents’, and by implementing these in some computational way, it should be possible to replicate the process of reading an ancient document. Second, the corpus of letter forms, plus the additional data regarding the language used at Vindolanda (Chapter 3) can provide the data to train such a system.

This chapter, and Appendix A, explore and describe the implementation of the artificial intelligence system which was developed to read the Vindolanda texts. An appropriate architecture, the GRAVA system (originally developed for aerial image understanding), was identified, adopted, and adapted for this research. Due to the diverse nature of this text's audience, this chapter provides an overview of the system for the less computer-oriented reader, whilst Appendix A describes the technicalities and mathematical underpinnings of the system in detail. Those readers without a background in computing or engineering science should find this chapter sufficient to explain the workings of the computer system. The results of the system are presented in Chapter 5.

The system in question, GRAVA, allows small modules, or computer 'agents', to work together to compare and contrast different types of information. It does this using minimum description length. This chapter introduces the GRAVA system, explains briefly the application of the original system, before demonstrating how the system could be applied to the data gathered from the Vindolanda texts. It is then explained how and why the GRAVA system had to be adapted to deal with the complexity of the Vindolanda texts, before discussing the basic underlying structure of the system.

4.1. INTRODUCTION TO THE GRAVA SYSTEM

The GRAVA (Grounded Reflective Adaptive Vision Architecture) system, developed by Dr Paul Robertson (2001) provides a way to implement different modules, or agents,¹ of computer programs

¹ Agents are anything that perceive their environment through sensors, and act upon that environment through effectors. In computer science, and specifically artificial intelligence, a software agent is a piece of autonomous, or semi-autonomous, proactive and reactive, computer software. Many individual communicative software agents may combine to form a multi-agent system. Agent cooperation can span semantic levels, allowing hierarchical systems to be constructed (see Russell and Norvig (1995) for an overview of artificial intelligence and agent-based systems). In the case of the GRAVA software, 'agents' are autonomous computer programs that match input to known models of data, and assess how well the unknown input data matches these models.

which can compare and pass different types of information to each other. These individual agents (programmed in Yolambda, a dialect of Lisp),² can be stacked up to represent a hierarchical system, such as McClelland and Rumelhart's 'Interactive Activation and Competition Model of Word Recognition' (1981) (see S. 2.8.2; Rumelhart and McClelland (1982); McClelland and Rumelhart 1986*b*), or the proposed model of how papyrologists read ancient texts (see S. 2.8.3). Building such agent-based systems in the field of artificial intelligence is not new (Huhns and Singh 1998), but the GRAVA system is unique in that it uses description length as a unifying mean of comparing and contrasting information as it is passed from one level of the system to another.

'Description length' is calculated from the probability of an input, such as the data describing the shape of an unknown character, matching a known model. By adding the description lengths of the outputs of different agents, it is possible to generate a global description length for the output of the whole system. Communication theory has demonstrated that the shortest description, or explanation, of a problem is the most efficient (Forster 1999; and also Ch. 1. n. 35). Therefore, the smallest global description length generated from many runs of a system is most likely to be the correct solution to the problem: the minimum description length or MDL. It is important to note here that the GRAVA system is not a recursive system in the way that the model proposed in 2.8.3 is, instead the GRAVA system propagates different solutions to the problem on each run of the system. The global description length of each of these runs is compared, and the smallest description length generated from any one of these runs, or minimum description length, is chosen as it would indicate the most likely solution to the interpretation problem. (Information regarding the background of MDL is provided in

² Yolambda was designed to be a compact dynamic language for developing software as well as a platform for experiments into language design and implementation. It is also used in reflective programming research. Yolambda was developed by Dr Paul Roberston, currently of MIT. It is a dialect (a variation) of Lisp, the favoured language of artificial intelligence research. See Laddaga and Robertson (1996) for more information on Yolambda, and Norvig (1991) for an introduction to Lisp and examples of its use in artificial intelligence.

S. 1.6.3. Further explanation of the generation of description length in the GRAVA system is provided throughout Appendix A.)

4.2. THE ORIGINAL SYSTEM DEMONSTRATED

The GRAVA system was successfully used to build a system that could effectively 'read' a handwritten phrase. This was done by having three levels of agents, the first which detected low-level features of text (such as junctions and the top and bottom endpoints of characters), the second which compared the data to a known database of character forms, and the third level which compared combinations of characters to a database containing words. (Further explanation of the architecture of the GRAVA system is provided in Appendix A.)

First, the system is trained on a known corpus of words, and information about the words, letters, and characteristics of letter forms contained in the corpus is calculated and stored, so that unknown input data can be compared to the known models. Unknown image data is then fed into the system. Robertson demonstrates how the low level agents identify features and generate a description of these features based on the tops, bottoms, and junctions of characters. The system compares these features with the models of characters that were computed earlier from the known corpus. This comparison generates description lengths for how well the unknown features match each of the known characters in the corpus. One of these characters, and its description length, is passed to the next level to make a sequence of characters: the character is chosen by a 'random weighted' selection process, which means that, whilst the character is selected at random, those with a higher probability of matching the known models are selected more often than those with a low probability of matching the models. (The locally most likely choice may not make sense at the global level: if only the most likely character on the feature level was chosen every time, the correct overall combination may never be found.) The system then compares the resulting strings of characters, or 'words', with known words in the corpus, generating description lengths for each. A global description length for the phrase is calculated by

summing the description lengths of each of the characters and words. In subsequent iterations different possibilities are generated, and the system searches for the best fit by looking for the configuration which gives the lowest global description length. This global minimum description length corresponds with the most probable 'reading' of the text, which in many cases is the correct reading.

In this way, Robertson shows that MDL formulation leads to the most probable interpretation, and also that global MDL mobilizes knowledge to address ambiguities. (In the second iteration of this example, fully half of the characters were 'guessed' wrongly by the system, but in the fifth iteration all ambiguities were resolved, because the correct choices were what led to the global minimum description length.) The system produces a different description length on each iteration, but in this implementation only the results which improve on the previous results are outputted, so that the best interpretation of the data becomes obvious to the user. Robertson's system provides a robust, easily implemented architecture that produces convincing results in a short time frame.

Robertson used this *MARY HAD A LITTLE LAMB* example to illustrate the power of his new architecture on a simple problem, before developing special purpose agents for aerial image understanding (the development of a self-adaptive architecture³ for image understanding being the focus of his thesis). However, this example provided a useful starting point for a system to read the stylus tablets as it provided the architecture to construct, develop, and adapt a system to read ancient texts in the manner described in the model of section. 2.8.3, albeit concentrating on the lower levels of the model: the feature, character, character set, and word levels. At this stage of the research no modelling of grammatical structures has been carried out (see S. 6.2), and given the complexities of generating a complete meaning of a document, it is better to concentrate on the elements of the process that can be used to speed up the historians' readings of the tablets: the identification of individual letter forms, likely combinations of them,

³ Self-adaptive software is designed to evaluate its own behaviour. When this evaluation suggests that it is not accomplishing its intentions, the software changes its behaviour to improve functionality or performance (see Robertson and Laddaga 2004, for a discussion regarding self adaptive software and how GRAVA has been used to explore this concept).

M A R Y H A D A
L I T T L E L A M B



M A R Y H A D A
L I T T L E L A M B



=> (runCycles #t)

Description Length = 306.979919

Description = (t=0 b=0 j=0 p=0 t=0 b=2 j=0 p=1 t=0 b=2 j=2 p=1
t=0 b=2 j=2 p=1 t=2 b=1 j=1 p=1 t=0 b=0 j=0 p=0
t=2 b=2 j=2 p=1 t=0 b=2 j=2 p=1 t=0 b=0 j=0 p=1
t=0 b=0 j=0 p=0 t=0 b=2 j=2 p=1 t=0 b=0 j=0 p=0
t=1 b=1 j=0 p=1 t=1 b=1 j=0 p=1 t=2 b=1 j=1 p=1
t=2 b=1 j=1 p=1 t=1 b=1 j=0 p=1 t=2 b=1 j=1 p=1
t=0 b=0 j=0 p=0 t=1 b=1 j=0 p=1 t=0 b=2 j=2 p=1
t=0 b=2 j=0 p=1 t=0 b=0 j=2 p=1)

Description Length = 116.690002

Description = (M A A E H A D A I L T E S T I R M B)

Description Length = 65.699996

Description = (M R A Y H A D R L I T T L E L A M B)

Description Length = 61.749996

Description = (M R A E H A D A L I T T L E L A M B)

Description Length = 41.649997

Description = (M A R Y H A D A L I T T L E L A M B)

Figure 4.1. This figure shows how the original GRAVA program takes in images of a text, recognizes features of the text, then resolves this into encoded models of letters which can be compared to other data. First, the input data is shown. The feature detection agent detects the top and bottom endpoints and junctions of the characters. Each character is described as a list of endpoints and junctions, for example the first letter A in the word

and words, as identified in Chapter 2. The use of such features as end points, and junctions in the GRAVA system also provides a useful introduction to the problem of reading the stylus tablets, as this information has already been captured in the data set (see S. 3.5), and so a first run of the system on the data was easy to implement.

4.3. PRELIMINARY EXPERIMENTS

First trials were done, utilizing the annotated ink tablet images, where end points of the characters were used to provide the identifying features which were passed upwards to the character level (in exactly the same way as the example above, Figure 4.1, but only utilizing end points as the source of character data instead of end points and junctions, for simplicity in this first implementation of the system).

First, model data were generated using the majority of the annotated images of the ink stylus texts (the test set, a section of tablet 255, being excluded from the training set to allow a fair comparison). Second, the annotated test section of 255, see Figure 3.14, was read into the system, and the system attempted to resolve the features into words within fifteen run cycles. Results were encouraging.

However, there is a significant limitation to this approach in using the data to 'read' unknown characters. Ultimately, this system aims to dovetail with the work done using phase congruency-based local energy measurement to identify possible candidate handwriting strokes within

Figure 4.1. (*Contd.*) MARY has no obvious top endpoint, two bottom endpoints, and two junctions, which means it is described as ' $t=0$ $b=2$ $j=2$ '. The description of each character is delimited with 'P' to describe that a character is detected at all. The combination of top and bottom endpoints and junctions for each individual character is then compared to the models of characters by the character agent, and one of these letters is passed up to the next level, the word agent, using a random weighted process to ensure that different combinations of characters are passed to the word agent. Combinations of characters are compared to words from the training corpus in the word agent (if no word is present which fits the data, the characters are left as a string of individual characters, separated by spaces). The description length for the characters and words in the phrase is summed to generate an overall description length for the output. The result with the shortest description length overall is the correct solution (Robertson 2001: 24.).



Figure 4.2. The screen output of the GRAVA system, showing the description length of different iterations. The text was correctly interpreted in the seventh iteration. The first section of output lists, for each stroke, a set of co-ordinates. The subsequent section outputs the summed description length for each resulting string (which is a possible solution to the problem). A space between characters indicates that character is being read individually, characters together are identified as words. The system only prints output on iterations where the MDL is shorter than that of the previous output(s). The number of run cycles can be chosen by the user: in this case it was 15 iterations. The output with the shortest MDL is the correct interpretation.

the stylus tablets (see S. 1.3, also Schenk 2001; Molton *et al.* (2003); Pan *et al.* (2004); Terras and Roberston (2004), Robertson *et al.* (forthcoming); Brady *et al.* (2005)). To investigate whether the system could cope with automatically generated representations of characters (or ‘computer’ annotated images as opposed to ‘hand’ annotated images), the same section of ink tablet 255 as above was analysed, utilizing the techniques developed for the analysis of the stylus tablets.

4.3.1. Automatically Annotating Stroke Data

The analysis⁴ had three stages, culminating in the generation of a representation of a character as a collection of strokes. First, features were detected (using phase congruency-based local energy measurement; see S. 1.3). Second, the image was segmented into separate sections. Finally, a stroke description was generated, which described each stroke by its location, shape outline, central line, end points, and junction

⁴ This analysis was carried out by Dr Xiao-Bo Pan, formerly of the Robots Group, Engineering Science Department, Oxford University, now at Mirada Solutions Ltd.

points. This description was output to an XML file, using the same markup as the GRAVA annotator program (see S. 3.6), allowing it to be read in and processed by the GRAVA architecture in the same way as a hand annotated image. Unfortunately, the resulting data is significantly different, in some ways, than that of a hand annotated image, making comparisons regarding end points (and/or junctions) alone worthless.

It can be seen that there are numerous occurrences in these two characters alone where the strokes and endpoints are identified completely differently. With the hand annotated U, the character is shown as one stroke with two endpoints, both in the upper part of the character. In the automatically recognized character U, the character is split into two large strokes, with one smaller stroke to the right also being included in the character. This gives a total of six endpoints. The S is also significantly different. The hand annotated S is shown as comprising of two strokes, one long, one short, with



Figure 4.3. A section of the annotated ink tablet 255 showing the characters 'u' and 's'. On the left is the image annotated by the automated process, on the right the hand annotated images. Rectangular areas denote junctions and endpoints. It is obvious that the endpoints annotated by the automatic process fall in different places to those annotated manually, especially in the case of the letter S.

a junction between them in the upper left of the character. The automatically recognized S also has two strokes, but the longer is at the top, and the shorter at the bottom. There is no junction, and there is a significant break between the two. Although both have four end points, two of these are in very different places.

This difference is hardly surprising, due to the noisy nature of the tablets, and the difficulties in automatically identifying features accurately within images (see Gonzalez and Woods (1993) for an introduction). It is often difficult for a human to segment overlapping areas in an image, never mind a computer program (witness the mistake made when annotating 'ussibus' by hand as 'ussibuss' due to the unclear nature of the final characters, see SS. 3.7.2, 4.3, and 5.2). However, this test indicates that the use of end points as the sole identifiers of characters in this application of the system would be unsuitable. End points turn out to be unstable, in that small changes in image quality can create large differences in the number and location of the end points. This instability makes end points a poor choice for character identification. The automatically generated annotations are compared to the character models built from the hand annotated characters, and so they have to hold a certain level of similarity for the system to function. The automatically annotated data generated from the test set was tried in the system, and it did not succeed in producing the correct reading. A more stable, graphical means of encoding and modelling the letter forms was needed, as a means of employing the available data in the feature level of the system. This would have to compare the most important data regarding the characters: their constituent strokes.

4.4. THE CONSTRUCTION OF CHARACTER MODELS

The database of annotated characters contains various information about the characters, some of the most important being the stroke data. It is possible to generate a model, or 'average' representation of each character from this data. This is done by collating the stroke data of each instance of the characters in the database onto a standardized grid, which results in a general representation of each type of

character. These are subsequently used as the models to which unknown characters are compared.

A further explanation of how character models were constructed is provided in section A.6, but it is enough to understand that the character level of the system was adapted to compare not the end points and junctions of the strokes, but the actual strokes of unknown characters to the stroke data contained within the character models. Full alphabets of the character models generated from both the annotated ink and stylus texts are given in section A.6.5, and these can be compared with individual stroke data in Appendix C.

4.5. THE NEW SYSTEM EXPLAINED

The new system is an adapted version of the original GRAVA system, where the character agent is more sophisticated, working with stroke data instead of end point data. All agents work with data generated from the Vindolanda corpus, as discussed in Chapter 3: the character agent uses the set of character models plus data regarding the frequency of letters in the Vindolanda corpus, whilst the word agent utilizes a list of words generated from the documents read at Vindolanda so far.

To input an image of an unknown document, this image has to be annotated (either manually or automatically, see SS. 3.7 and 4.3.1) to identify the key features, and the stroke data prepared for comparison (see A.6 for more detail). The stroke data is then passed on to the



Figure 4.4. The standard representation of the character M given in Bowman and Thomas (1983). The character model of the character M generated from all the ink characters in the training corpus. The character model of the character M generated from all the stylus characters in the training corpus.

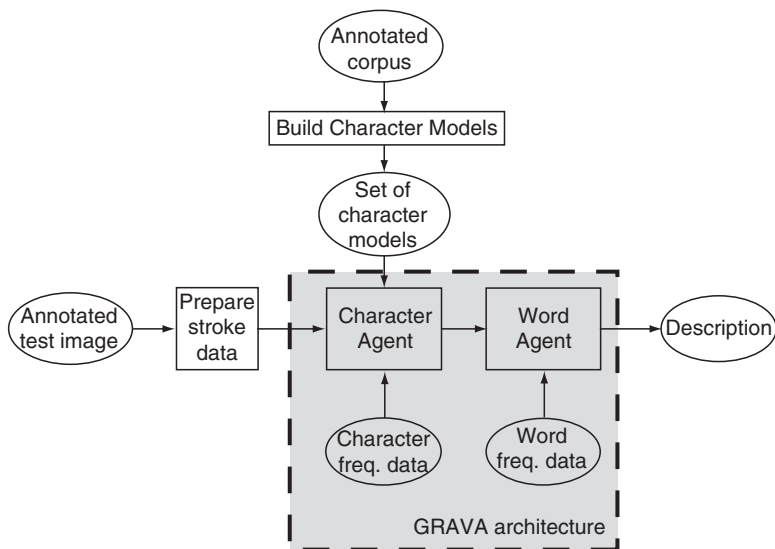


Figure 4.5. Basic schematic of system. Robertson's GRAVA architecture is highlighted to indicate the processes which are carried out as part of the final run of the system.

character agent, which compares this data to the character models (generated previously from the annotated corpus), and calculates a description length for the match of this unknown data to each character in the training corpus, and the frequency of characters which appear in the corpus ('i' is much more likely to appear than 'q', for example). One of these characters is selected by a random process, and is passed, with its description length, to the word agent. As characters are passed to the word agent, they form a string of characters, these resulting 'words' are then compared to known words in the corpus, and a description length is calculated for how well they fit known words from Vindolanda. The description length from the character agent and the word agent is then added, giving a global description length for that run of the system. Subsequent runs of the system generate different global description lengths, as the random process in selecting which letters are passed to the word agent assures that different combinations of characters that may fit are put forward as possible interpretations of the problem. When the description length of these

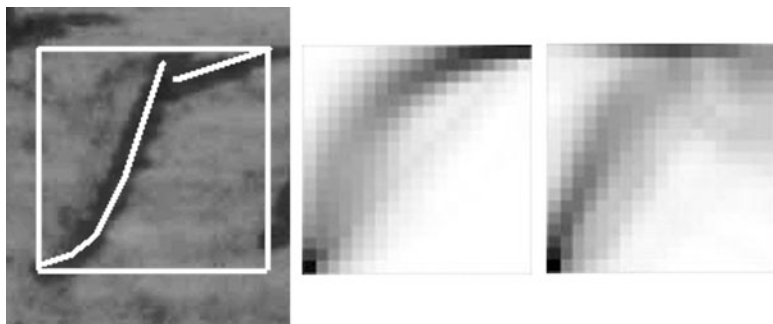


Figure 4.6. Letter S from ink tablet 255, compared to all the ink models in the training corpus is most likely to be S (left) or R (right).

separate runs of the system are compared, the minimum description length gives the overall most likely interpretation of the test image.

4.6. AN AGENT IN ACTION

This process can be demonstrated by showing the results given when an unknown character (the second S from ‘ussibus’, above), is compared to all of the available models generated from the ink tablets. In this case, the models most likely to fit the character were identified as S and R, shown graphically in Figure 4.6.

The strokes were identified in the unknown character and compared to all available models, generating a description length for how well the unknown character matches the known models for each model character. This is added to the description length generated from the probability of that letter occurring (computed from the lexicostatistics), to give the overall description length of the letter. In this case, it can be clearly seen that the letter is most likely to be S or R (those results highlighted with a dark strip for clarity): the shorter the description length, the more likely the identification is.

The last line in Figure 4.7 indicates the result of the match. The character data has been estimated to be an S with a description length of 82.009643 (which gives a probability of 0.858364, as description

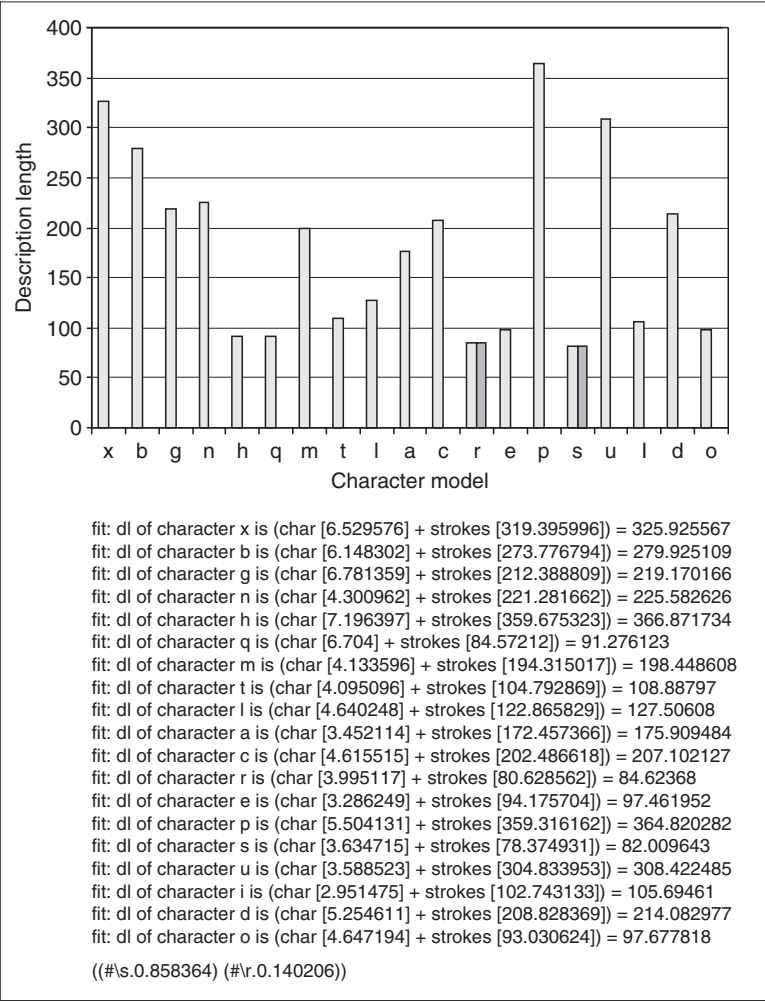


Figure 4.7. Fitting character models to the letter S: comparing an unknown to model data, based on the stroke data and character frequency data from the training corpus. The order of the character models in the bar chart is of no significance. The textual portion shows the text output from the model fitting program in debug mode. It can be seen that the input is matched to every model in the training corpus, and a description length calculated for each match.

length is the negative log of probability to the base 2, see S. A.2), or an R with a description length of 84.62368, which is a probability 0.140206. The remaining probability of 0.00143 is divided out among the other characters (probability always adds up to 1), which means that the probability of those occurring (even the nearest contenders, H and Q) is so small as to not be worth considering. A 'random weighted' statistical sampling method (see SS. A.2.2 and A.7) is then used to choose one of the most likely identifications, R or S, to pass up to the next level: the word agent.

The word agent works in a very similar fashion, comparing the string of characters delivered from the character agent to the list of known words in the corpus, and selecting one of the best fits as the final interpretation. Further information regarding the word agent can be found in section A.8.

The description lengths propagated by the different agents are then combined, to give an overall description length for the words identified in the text. The minimum description length gives the overall most likely interpretation of the test image in relation to the data held in the training corpus.

4.7. CONCLUSION

This chapter has provided an overview of the genesis and development of the system used to propagate interpretations of the Vindolanda texts. The concept of minimum description length has been introduced and demonstrated, and the basic architecture of the system has been explained. Further technical information is provided in Appendix A. But while it is all very well saying that minimum description length should provide the most likely output of what is written in theory, how does it work in practice? Chapter 5 shows the results of the system as applied to the Vindolanda texts, whilst Chapter 6 discusses how the system should be developed to enable more texts from Vindolanda to be read.

Results

Co-authored with Dr Paul Robertson

‘Now—here we go!’ He reached up and pulled a switch on the panel. Immediately, the room was filled with a loud humming noise, and a crackling of electric sparks, and the jingle of many, tiny, quickly-moving levers;...and sheets of quarto paper began sliding out from a slot to the right of the control panel...They grabbed the sheets and began to read. The first one they picked up started as follows ‘Aifkjmbasaoegweztppln-voqudskigt&—fuhpekanvbertyuio, lkjhgfdsazxcvbnm, peru, itrehdjkg mvnb, wmsuy...’ They all looked at the others. The style was roughly similar in all of them. Mr Bohlen began to shout. The younger man tried to calm him down.

‘It’s all right, sir. Really it is. It only needs a little adjustment. We’ve got a connection wrong somewhere, that’s all. You must remember, Mr Bohlen, there’s over a million feet of wiring in this room. You can’t expect everything to be right first time.’

‘It’ll never work,’ Mr Bohlen said.

(Dahl 1997: 24)

The research so far has developed a model of how experts read ancient texts, which refines this process into defined units that can be implemented in a computer architecture. A computer system—GRAVA—was identified as the means to computationally construct the model. However, the original GRAVA system had to be developed to deal with the complexity of information contained within the Vindolanda texts, and a corpora of letter forms, character frequencies, and words had to be constructed to provide data which would

allow the system to ‘reason’ about information contained within the Vindolanda texts.

Once built, the adapted version of the GRAVA system was used to analyse various sets of test data, to see how effective it could be in producing the correct ‘reading’ of a text. First, a section of tablet 255 was analysed, using the character models ascertained from the ink tablet corpus as the basis of comparison. This gave encouraging results, and also shows the asymptotic nature of the system’s convergence on a solution. This experiment was then repeated with the same section of tablet, which in this case had been annotated in a different (wrong) manner, to see how the system coped with more difficult data. A section of stylus tablet was then analysed, using first the set of character models derived from the ink corpus, and secondly the set of character models derived from the stylus corpus, to indicate how successfully the system operates on the stylus texts. Finally, a section of ink tablet which had been automatically annotated (using the techniques described in S. 4.3.1) was analysed, to indicate if this implementation of the system provided a possible solution to the problem of incorporating data generated from the image-processing algorithms into a knowledge-based system.

Although a complete stand-alone system has not yet been developed, these results demonstrate a proof of concept: that MDL can be used to compare and contrast semantically different information, and propagate realistic interpretations of image data within a reasonable time frame. Therefore, this chapter indicates that this technique could form the basis of a future stand-alone system to read further examples of the Vindolanda texts, and suggestions regarding future work which could improve the success of the system are presented.

5.1. USING INK DATA FOR INK TABLETS

The first complete run of the adapted GRAVA system was carried out on the section of tablet 255 previously used as a test set for the system (as in S. 4.3, although in this case the updated, correctly annotated version was used). The set of character models used in this run was that derived from the ink tablet training corpus. The test data section



Figure 5.1. A section of ink tablet 255, annotated with ‘ussibus puerorum meorum’ (see Figure 3.14) within the GRAVA annotator program (see 3.6). This section is labelled 255front7, so-called as it is the seventh line of text on the front of tablet 255.

was input into the system, 25 iterations were specified, and the output of the system is shown in Figure 5.2.

In this run of the system, eight iterations took place before the correct answer was generated. The system carried out 25 iterations on these data, but the minimum description length generated occurred in iteration 8, and so these are the last data shown (the system carried out a further 17 iterations, the results of which are not shown as only results which improve on the previous proposed solutions are outputted). It should also be stressed that, with all of these results, the experiment was run a number of times, and the results presented are the best case, where the system generates the correct result in the fewest iterations.) Previous outputs from iteration 1, 2, 3, and 5 had been possible solutions, but that from iteration 8 proved the best, given the data provided. This was also the correct solution. The GRAVA system successfully reconstructs the correct reading of the text in a short time, on this occasion.

5.1.1. System Performance

Although, above, the correct output was generated in merely 8 iterations, because of the stochastic nature of the process there is a possibility that the correct answer may never be found. If it is generated (in practice the correct output is usually generated within a few iterations) the number of iterations taken to reach this answer will be different on each run. This can be shown by determining the

```

Grounded Reflective Agent Vision Architecture (GRAVA) Version 2.0
Yolambda listener pushed. Type: exit to return to GRAVA

=> ... load the system and the data ...

=> (run Cycles 25)
iteration 0 DL=440.220794 interpretation = ( ... ((2482 252) (2517 250))) ... )
iteration 1 DL=64.085075 interpretation = ( u r s i b u s  puerorum m n o a u m )
iteration 2 DL=49.374412 interpretation = (ussibus puerorum m n o r u m )
iteration 3 DL=48.831413 interpretation = (ussibus puerorum m n o a u n )
iteration 5 DL=47.816696 interpretation = (ussibus puerorum m e o a u m )
iteration 8 DL=36.863136 interpretation = (ussibus puerorum meorum )
iteration 25

=> : exit

```

Figure 5.2. Output from a successful run on the section of 255, using the ink tablet character models. This output was taken from the screenshot of the system (as presented in S. 4.3) to make it easier to read. First, the system loads, and the data is inputted. The number of run cycles is specified: this can be any number, but the higher the number of iterations, the more likely the solution will be found. The first section of output, which lists, for each stroke, a set of co-ordinates, is omitted: we have opted to show the interesting section of interpretation of the data. The main section shows the output of the summed description length for each resulting string (which is a possible solution to the problem). A space between characters indicates that character is being read individually, characters together are identified as words. The system only prints output on iterations where the MDL is shorter than that of the previous output(s). As the system converges on a solution, the description length becomes shorter. For example, a description length of 49.37 is given for the interpretation ‘ussibus puerorum m n o r u m’, whilst the interpretation ‘ussibus pueorum m e o a u m’ gives a shorter description length of 47.81, as the combination of characters in the last word is more probable when compared to the statistics contained within the bigraph analysis of the language and the match of the unknown characters to character models. The output with the shortest MDL is the correct interpretation.

average description length that is generated over a variety of runs on the same data. The example, 225front7, as used above, was run 200 times, with 25 iterations specified in each run. By plotting the average description length generated from each of the 25 iterations over the

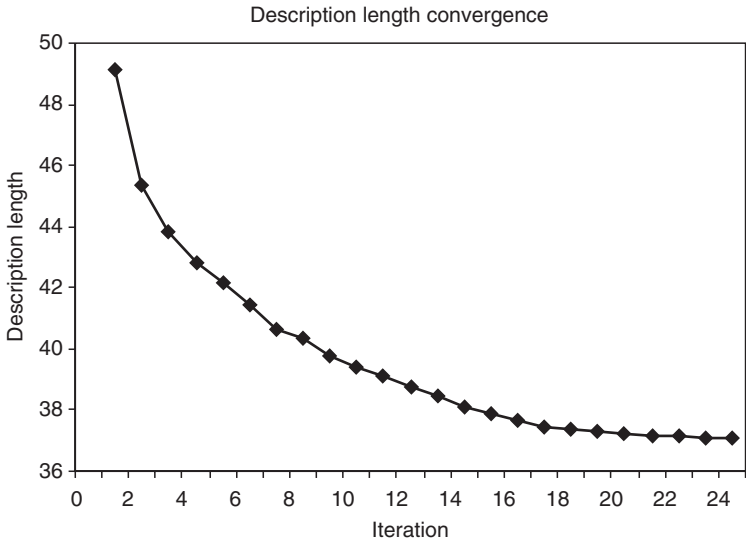


Figure 5.3. Description length convergence over iterations. The system was run 200 times on the same piece of data, with 25 iterations on each run. The average description length for each iteration is plotted, showing the asymptotic nature of the algorithm. It shows, for this example, that a few iterations of the system would not be enough to generate the correct interpretation: this would occur around the 20th iteration (or thereabouts) on the average run. However, due to the random nature of the Monte Carlo algorithm used, it may very well occur on the first run, or the last, or not at all.

200 runs, it becomes obvious that the system converges on a result.¹ The MDL of the ‘correct’ result in this case was 36.863143, whilst after 25 iterations the system averaged 37.08221. This is due to the fact that the average asymptotically approaches the perfect value because of the Monte Carlo sampling methods utilized (see S. A.1.2).

¹ Many outputs of computer systems generate a sequence of approximate solutions; if this sequence tends more and more closely to the true solution of the problem, the algorithm is ‘convergent’. Mathematical analysis can demonstrate the rate of convergence to the known solution across various iterations, and thus the efficiency of the system. See Zhang (2004) for an introduction to convergence analysis.

Of course, this example only pertains to the section of 255 used as the test set. The more complex the input data, the longer the system will take to converge on a solution (which will have a different MDL). This example, however, demonstrates that the system is effective at generating likely solutions within a relatively short time frame, and that solutions are quite likely to be found.

5.2. USING INK MODELS FOR AN UNKNOWN PHRASE

A second section of ink tablet was analysed, this time to see how the system coped with a more confusing section of strokes, and also how it could identify a word that was not in the corpus. The first version of the annotation of 255front7 was used for this: where it had been unclear where one letter ended and another began (see SS. 3.7.2 and 4.3). The difference between the two versions is that the first was annotated USSIBUS, whilst the eventual, correct, annotation is USSIBUS. (The change was made when cross-referencing the various information ascertained from the papyrologists, and this example kept to see how the system coped with a more difficult segmentation problem.) Steps were taken to ensure that the word fragment USSIBUSS did not occur in the word corpus for this iteration, to see how the GRAVA system coped with this difficult input.



Figure 5.4. On the left, the section of 255 correctly annotated as US. On the right, the same section annotated incorrectly as USS, which was kept for test purposes. On the left, the letter u contains all the strokes, whilst these have been separated into two distinct characters on the right, inserting an extra letter ‘s’ onto the end of USSIBUS.

```
Grounded Reflective Agent Vision Architecture (GRAVA) Version 2.0
Yolambda listener pushed. Type: exit to return to GRAVA
```

```
=> . . . load the system and the data . . .
```

```
=> (run Cycles 200)
```

```
iteration 0 DL=440.220794 interpretation = ( . . . ((2482252) (2517250))) . . . )
iteration 1 DL=69.437759 interpretation = (u r s i b l n s puerorum m n o a u m)
iteration 3 DL=69.077354 interpretation = (u s s i b l n s puerorum m n o a u m)
iteration 5 DL=48.062644 interpretation = (u s s i b l n s puerorum m e o a u m)
iteration 6 DL=57.469482 interpretation = (u r s i b l n s puerorum meorum)
iteration 18 DL=57.109081 interpretation = (u s s i b l n s puerorum meorum)
iteration 25 DL=56.903217 interpretation = (u s s i b l t s puerorum meorum)
iteration 199
#f
=>
```

Figure 5.5. Output from the system, analysis of awkwardly annotated segment of 255, using the ink tablet character models. The system was set to run 200 iterations, to give enough chances for the correct interpretation to be generated on this complex example. The system manages to resolve the last two words of the phrase, but is unable to confirm a word identification of the first word: the characters are still treated as individual characters rather than part of a word string, as a matching word is not located in the word corpus. Nevertheless, the USSIB**S approximation generated would be enough to suggest to an expert the most probable interpretation of the strokes: the combination of characters presented here being the most likely according to their match of character models and the expected frequency of letters and combinations in the corpus.

These results are interesting on a number of levels. First, the system is confused by the last few characters in USSIBUSS, indicating that it is having problems identifying them. The first problem character is identified (not unreasonably) as an L, the second as either an N or a T. This shows how an unclear character can be assigned a number of possible solutions. Second, although the sequence of characters (USSIBUSS) is not in the word corpus, the system does a good job at reconstructing a possible string of characters, resulting in USSIB**S. This is partly due to the use of the character models, and also the use of letter frequencies and bi-graph frequencies. This approximate solution should be enough to give some indication to a human user of what the correct word may be (the system will

eventually have to interact with experts in this manner, see Chapter 6). Finally, the MDL generated from this solution to the problem is 56.903217. The MDL generated from the alternative (correct) annotation of the characters in section 5.1 was 36.863143. This shows that the most likely solution to identifying the letters will have the lowest MDL, and also that there is some need, in the future, to encapsulate the opportunity to reannotate difficult characters in a way that will eventually produce the lowest MDL to generate possible solutions.

5.3. USING INK MODELS FOR STYLUS TABLETS

It was suggested in section 3.1.2 that the letter forms from the ink tablets should correspond to those from the stylus tablets closely enough to allow models from the ink tablets to aid in readings of the letters of the stylus tablets, and these similarities were discussed and demonstrated in section 3.11. In this experiment, a section of stylus tablet 797 was analysed, first using models derived from the ink tablets, and in the subsequent section using the small set of models derived from the stylus tablet corpus. This section of tablet contained fairly common words, NUNC QUID (although it had taken the experts a substantial length of time to come up with this reading).



Figure 5.6. Section of stylus tablet 797, annotations showing NUNC QUID. See Figure 2.3 for a translation. This stylus tablet had been read by the experts and the contents known, allowing the testing of this system on the data.

Grounded Reflective Agent Vision Architecture (GRAVA) Version 2.0.
Yolambda listener pushed. Type: exit to return to GRAVA.

=> . . . load the system and the data . . . (Ink Models)

=> (run Cycles 200 #t)

iteration 0 DL=34.567626 interpretation = (n n n c d u i d)

iteration 1 DL=31.070133 interpretation = (n n u i d u i m)

iteration 2 DL=31.070131 interpretation = (n u u i d n i m)

iteration 5 DL=22.481197 interpretation = (u u u i q u i d)

iteration 82 DL=21.051862 interpretation = (n u n c q u i d)

iteration 199

#f

=>

Figure 5.7. Output of GRAVA system, section of 797 utilizing ink character models. Again, 200 iterations of the system took place to give enough chance that the best solution to the problem be propagated. The first few iterations indicate that the system was unable to identify the letters correctly, meaning no words in the database matched the letter sequence, but after 82 iterations the correct interpretation was found, with no subsequently better interpretation being generated by the remaining cycles.

Although a fairly small sample of text, it contains a few difficult characters (the curvy letter D, for example, and the large C): the results demonstrating how the system will cope with the differences in character forms between the ink and the stylus texts.

The system took 82 iterations to reach the correct interpretation, quite a high number of run cycles, probably due to the differences in forms between the character sets: the ink text models are not exact matches for the characters found on the stylus texts. However, their approximate similarities allowed the correct solution to eventually be found. The first few iterations, as suspected, show that the system had problems with the letter D, interpreting it as an M, and the large letter C, interpreting it as an I. However, the correct reading was eventually generated when enough run cycles were allowed to sort through the various hypotheses thrown up by the data.

5.4. USING STYLUS MODELS FOR STYLUS TABLETS

The same section of tablet 797 was analysed, this time using the small selection of character models generated from the annotated stylus tablets. This run of the system identified the correct response in only 18 iterations, making it much more successful than the run, above, where data from the ink character models were used. There are a number of reasons why this is the case. First, the character models generated from the ink and stylus tablets are *slightly* different, and this small difference must have a large effect on the comparisons. Secondly, although the stylus models character set is impoverished due to the lack of available data (see S. A.6.5), it contains almost all of the characters based in this sample (save for the letter U. The model for the character U was borrowed from the ink tablet models in order to be able to try this test. The letter U seems to be preferred by the recognizer over other characters in this run. This is probably because the model for U borrowed from the ink tablet models was based on a large number of samples whereas the models for the stylus characters were based on a small sample of characters). However, the stylus character model set

Grounded Reflective Agent Vision Architecture (GRAVA) Version 2.0.
Yolambda listener pushed. Type: exit to return to GRAVA.

=> . . . load the system and the data . . . (Stylus Models)

=> (run Cycles 25 #t)

iteration 0 DL=35.304573 interpretation = (u u n c q n i d)

iteration 1 DL=34.447456 interpretation = (u u u c q u l m)

iteration 5 DL=30.377746 interpretation = (n u n c q n i m)

iteration 12 DL=24.857675 interpretation = (u u n c q u i d)

iteration 16 DL=24.145236 interpretation = (u u u c q u i d)

iteration 18 DL=21.051862 interpretation = (n u n c q u i d)

iteration 24

#f

=>

Figure 5.8. Output from section of 797, utilizing the stylus character models set. In this case, 25 iterations of the system took place, to see if the solution could be found quickly, and the correct solution was propagated after only 18 iterations.

does not contain all the available characters, which would make it difficult to identify further examples where less common letters predominantly feature. Although comparisons with the stylus character models appear better than the ink character models, there is still a place for the ink character models. Future implementations of the system could have more than one model for each type of letter, as discussed below (S. 5.6).

5.5. ANALYSING AUTOMATED DATA

It has been shown, above, that the system can correctly interpret annotated images from the corpus which were generated by hand. This is all well and good, but in essence the system needs to work effectively alongside the image-processing algorithms that have been developed in tandem to this research (SS. 1.3 and 4.3.1). The section of 255 used as test data was analysed and annotated automatically.² This was then used in the system as test data. Due to the image data of the third word being incomprehensibly faint, these were elided from this test. The results are presented in Figure 5.10.

Although the system initially finds the correct result, it goes on to find a ‘better’ result that was wrong. This is because in the second word it identifies **ER*RUM, where* is an error. Eventually, the Monte Carlo search finds characters that match SOLEARUM well enough to get a lower description length than PUERORUM. However, the system did get the first word 100 per cent correct at the character level, and the



Figure 5.9. Section of 255 analysed using the automatic feature extraction algorithms. More information regarding the algorithms used to propagate this data can be found in Robertson *et al.* (forthcoming).

² Again, this was carried out by Dr Xiabo Pan, formerly of the Robots Research Group, University of Oxford, now at Mirada Solutions Ltd.

```

Grounded Reflective Agent Vision Architecture (GRAVA) Version 2.0.
Yolambda listener pushed. Type: exit to return to GRAVA.

=> . . . load the system and the data . . . (Stylus Models)

=> (run Cycles 200 #t)
iteration 0 DL=24 . 575424 interpretation = (  ussibus puerorum  )
iteration 1 DL=23 . 990461 interpretation = (  ussibus solearum  )
iteration 199 #f

=>

```

Figure 5.10. Output from an analysis of 255, using automatically annotated images and the character models from the ink tablets. The system went through 200 iterations, and propagated two solutions to the data, arriving at the correct interpretation on the first attempt, then finding another (incorrect) solution which fitted the models better!

second word was almost recovered despite a rather poor interpretation at the character level. This shows promise towards generating possible interpretations of images through utilizing this architecture. Further work needs to be done on the stroke extraction algorithms used in the automated annotation process, to encapsulate stroke data more fluidly, to enable this to be utilized by the character agent. The current algorithms have problems with the grouping of strokes, and identifying the continuation of strokes where the traces of them are faint. When these problems are overcome, and the automated output is in a clearer format, it will be easier to generate possible interpretations of the images using the GRAVA system. Work continues at the Department of Engineering Science at the University of Oxford to improve these algorithms. The results presented here show promise towards a possible solution to the problem of how to segue the image-processing algorithms into an information-based system to aid the papyrologists in the reading of the texts.

5.6. FUTURE WORK

This research has demonstrated how an MDL architecture such as the GRAVA system can be appropriated to generate possible solutions to interpretation problems encountered whilst trying to read

the ink and stylus tablets. Although the examples shown were fairly straightforward, and propagated interpretations of texts which had already been read, they indicate that attempting further analysis of the stylus tablets in this way will be a worthwhile endeavour. However, much can be done to improve the system, or to try and incorporate other types of information into the architecture to improve its functionality, before using the system to propagate readings of texts which have not been read previously (which, after all, is the absolute aim of this research).

First, it was shown that the stylus tablets were 'read' easiest when the stroke data was compared with the stylus tablet models, but that some character forms were not contained in the stylus character models data set. It would be easy to amalgamate the two sets of character models into one: not combining the character models themselves, but just adding the different character forms to the existing set, so that there were two (or more) forms of different characters in the model base. (It does not matter to the word agent if there are two different forms of the letter X at the character level, as long as one character gets passed up.) Also, the different letter forms of Q and D could be separated (where their descenders go right to left, or left to right), so that there were two character models of this letter present in the model set. By expanding and refining the character model set in this manner, the identification of unknown characters by the character agent should improve.

Second, there is a great deal of information regarding the character forms that was captured in the annotation process (S. 3.5) which has not been utilized in this system. For example, it has been shown that one of the most obvious features that experts look for when identifying letters were obvious ascenders and descenders, where strokes extend above and below the line. This information was captured in the annotation of the ink and stylus tablets images but has not been put to use in this system. It should be easy to incorporate further information into the character agent, to see if this aids in the recognition of unknown characters. There is a possibility that additional information would only affect the output of the character agent slightly, but it would be worth experimenting to see if it could be integrated into the system in this manner.

Additionally, there is a great deal of work which could be done to expand the database that the word agent uses. The current word list is

derived from the corpus of the Vindolanda texts exactly as it stands, with no changes made, or new words generated from other words in the corpus. Developing a system to output other possible word lists would be possible, but a major undertaking. This is covered in Chapter 6.

Finally, this research has indicated a proof of concept: that utilizing an MDL architecture can provide the necessary infrastructure as a basis to implement a system that can read and interpret images of the Vindolanda texts, and in particular, the stylus tablets. However, there is a long way to go before an integrated desktop application has been developed which will be able to aid the papyrologists in their day-to-day tasks. This depends on further development of this system, and further investigation into the image-processing algorithms to allow integration with this process. Again, this is covered more fully in Chapter 6, and work continues to turn this system into a working desktop application.

5.7. CONCLUSION

This chapter has successfully shown that an MDL architecture such as the GRAVA system provides the infrastructure to model the reasoning process the papyrologists go through when reading an ancient document. This provides the means to implement a system that works in a similar manner as the experts, generating possible, reasonable, solutions to interpretation problems. Although not a development of a complete stand-alone application, this research provides the means by which to go on and develop such a package which the experts can use to aid them in their task of reading the stylus tablets. It has been successfully demonstrated that information from the Vindolanda corpus regarding character forms and words can be used to generate plausible readings of tablets. It has also been shown that the development of the character agent that relies on data regarding the character strokes, rather than end points and junctions, is more effective at producing possible identifications, and that this character agent can more easily work with test information presented by the image-processing algorithms.

Although much more work needs to be done on both this system and the image-processing techniques before a desktop application can be implemented, this research has provided the first steps towards that aim. A possible method has been identified, implemented, and demonstrated, which enables the amalgamation of the image-processing and procedural information into a cohesive whole.

6

Conclusion

The king cried aloud to bring in the astrologers, the Chāl-dē-āns, and the soothsayers. And the king spake, and said to the wise men of Babylon, Whosoever shall read this writing, and shew me the interpretation thereof, shall be clothed with scarlet.

(Daniel 5: 7)

This research has demonstrated that an MDL architecture such as the GRAVA system can be appropriated to generate possible solutions to interpretation problems which span different semantic levels, and incorporate different types of data. In doing so, a system has been constructed which can read ancient documents in a similar way to papyrologists: the system described has an emergent behaviour similar to that of human experts and generates possible, reasonable, interpretations of texts, to aid the experts in their task. Although the examples shown in Chapter 5 were fairly straightforward, and indicate that the system can read texts which are already known, this indicates that attempting further analysis of the stylus tablets in this way will be a worthwhile endeavour. Research continues in the Department of Engineering Science, University of Oxford, to develop this system into a stand-alone package which can be used to propagate likely interpretations of the stylus texts. Much can be done to improve the system described here, and to incorporate other types of information into the architecture to improve its functionality.

Due to the interdisciplinary nature of this research, many topics have been covered, however briefly, in this monograph. There is scope for further research in almost every facet. Considerations of future work will focus on two main areas: other possible approaches

to knowledge elicitation to enable further understanding of how experts read ancient documents, and the enhancement and development of the MDL-based GRAVA system described in Chapters 4, 5, and Appendix A to increase its accuracy, and eventually deliver an application to the papyrologists.

This chapter also highlights the overall contribution the research has made to its many constituent fields. Suggestions for how this type of system could be adopted to aid humans in the interpretation of other types of data are made, and an evaluation of the research is presented: the cognitive visual architecture developed being a testament to the value of interdisciplinary research.

6.1. KNOWLEDGE ELICITATION

Although the study undertaken in Chapters 2 and 3 considering how experts read ancient texts and identify individual letter forms was as comprehensive as time and facilities would allow, there remains a considerable amount of research that could be undertaken in this area. First, using the techniques applied in Chapter 2, it would be useful to look at the techniques used by experts who deal with other linguistic systems. This would indicate whether the findings of this research can be applied to the field in general or whether they merely relate to the reading of the Vindolanda texts (and thus, whether this computing package could be applied to reading other types of documents, or merely the Vindolanda texts). Secondly, although the data set constructed in Chapter 3 is, again, comprehensive, there are a few more steps that could be taken to understand the palaeography of Vindolanda, and to improve the quality of the data set. Finally, little work has been done on how the experts utilize visual clues to build up an understanding of the texts; the research in this monograph mainly deals with verbal and written evidence, and does not investigate the type and quantity of data the experts visually study whilst trying to read a text. By carrying out a series of tests to investigate oculomotor action, it may be possible to see which features of the texts are the most important, and problematic to identify, when trying to piece together a reading of ancient documents.

6.1.1. Reading Ancient Texts

This research has focused exclusively on how experts read the Vindolanda ink and stylus tablets. It would be possible to easily extend the remit of the survey, utilizing the techniques used in this book, to investigate the readings of different types of ancient text. This could include such texts as papyri, ostraca, inscriptions, and lead tablets from similar and different periods to that of the Vindolanda texts. Similarly, the reading of texts that contain different letter forms, such as Capitalis and New Roman Cursive, as well as different linguistic systems, such as Cuneiform, could be investigated. Researching the work of experts who read and translate bilingual texts may also be fruitful in detailing how such a task is carried out (see Boswinkel and Pestman 1990; Adams *et al.* 2002, for an introduction to bilingual texts). A comprehensive study would contribute greatly to the literature available regarding how humans read in languages other than their own native tongue, and also how experts cope with ambiguities in texts. Although it is suspected that the findings of the research contained within this monograph would be applicable to other areas, a larger scale investigation would show whether this was the case. Such a study would be time-consuming and labour-intensive, and depend on the cooperation of experts from related fields, but could prove to be an excellent research topic.

6.1.2. Collecting Further Data

As much data as possible concerning how the experts approach and reason about ancient texts were collected given the time and resources at hand. However, if they were asked to discuss their reading of further documents from Vindolanda, and the resulting transcripts were analysed in the same manner as in Chapter 2, the results presented in this monograph could be statistically verified. This would also give a larger sample set from which to generate statistics regarding the relationships between different types of information used in the discussions regarding the Vindolanda texts (see S. 2.5.2.1). The two ink texts that were used in the initial experiment were chosen because some sections were easily readable and some were more abraded. It would be possible to set the experts different

tasks using more abraded images to see how that affected the type and amount of information they discussed regarding the documents. Again, this would depend on the experts themselves being willing to commit time and permission to the project.

6.1.3. Collecting Character Information

The data gathered regarding the type of information used when identifying characters in Chapter 3 was more than enough to be able to construct a reliable data set of annotated images. However, it may be useful, at some stage, to use a knowledge elicitation tool such as WebGrid (see S. 3.2.3) with the experts, to directly capture information regarding each individual letter form, and each letter's characteristics. This could provide another substantial source of information that could be used as an alternative data set with which to compare unknown letter forms, and also highlight any areas not covered by the knowledge engineer. Use of this program by more than one expert would also allow a direct comparison of the information each individual utilizes (the program provides statistical tools to analyse the different types of information entered by different users). This could provide further data for Chapter 2, showing how individuals make use of similar or different types of information when reading an ancient text.

The character corpus built up by annotating the ink images provides a representative sample of the characters found within the ink documents. However, the coverage of the character forms found on the stylus tablets was poor, due to the small sample size of these texts. This raises an interesting opportunity. As more stylus tablets are read, images of them can be annotated and added to the corpus, thus improving the training set on which to base further readings. Character models will therefore be refined as more documents are read and their characters added to the existing corpus: making the feedback mechanism the papyrologists instinctively use an explicit part of the computer system.¹ Also, in the future, when a document has been successfully 'read' using the system, other data regarding

¹ As papyrologists learn more about the letter forms and words used in certain types of documents, the accuracy and speed at which they can read them improves, see Bowman and Thomas for an example (2004).

that document, for example words, can be fed back into the data set, thus increasing its relevance, and improving system performance.

As mentioned in section 3.12, to check the quality of the annotations made in the corpus, it would be prudent to undertake a further program of annotation, where the manually added codes could be reannotated, and the resulting data automatically compared to the initial annotations. Any differences in the data would then be easily detectable.

6.1.4. Studying Oculomotor Action

The majority of the knowledge elicitation exercises carried out during this research were based on the collection and analysis of textual data. There was no investigation done into how the experts read the ancient texts on a physical level, to see which features of the text they visually concentrated on, or, alternatively, features they easily identified and did not have to focus on. One way to investigate this would be to study the pattern of eye movements used by the experts as they attempt to read an ancient text. This type of research would have been unfeasible until recently, as oculomotor technology has only just become cheaply and readily available.

There is a widely held assumption that there is a close correlation between pattern of eye movements and mental processes undertaken (Liversedge *et al.* 1998). However, the relationship between vision, perception, reading, and language is still unknown (Gregory 1994: 204). The study of oculomotor action is critical for the efficient and timely acquisition of visual information regarding complex visual-cognitive tasks, such as reading. How humans acquire, represent, and store information about the visual environment is a critical question in the study of perception and cognition, and data regarding eye movements provides an unobtrusive measure of visual and cognitive information processing (Henderson and Hollingworth 1998). How we move our eyes when we look at a picture depends on what perceptual judgement we are asked to make (Yarbus 1967). There is a growing body of literature about the role of eye movement in scene perception, and that which exists suggests that informative areas receive more fixations than others (Underwood and Radach 1998).

The role of eye movement in reading has been the focus of many experimental studies in the last twenty years (Rayner 1998), including the role of eye movement in reading music (Furneaux and Land, 1997). Techniques developed could be used to analyse how experts read ancient texts. Various oculomotor measures, such as the duration of fixations (where the gaze is concentrated) and saccades (the brief, rapid movement of the eye from one position of rest to another) whilst reading a document, could be undertaken. Although untangling the process of reading a document from start to finish from the resulting data would be a complex task, the results would show the features of a text the experts have most difficulty identifying, or are most important for the identification of individual characters. This, in turn, could affect the focus of the work that is being carried out on feature detection and image processing, as it would indicate the areas of importance, and confusion, for the experts when reading an ancient text.

6.2. EXPANDING THE EXISTING SYSTEM

Although there remains a lot of research that could be done regarding understanding how experts read ancient texts, the main focus of further work should be towards expanding the system; to increase the amount and type of data it relies on, to improve its functionality, and to implement a stand-alone desktop application which can be routinely used by those attempting to read the Vindolanda stylus texts. There are a number of ways in which improvements could be accomplished. First, more statistics could be generated from the Vindolanda textual corpus, and these integrated into the system. Secondly, the word list generated from the textual corpus could be expanded by analysing and developing the existing corpus. Further linguistic work could be undertaken on a grammatical and semantic level. Character information that was captured during the annotation process could be integrated with the existing system, which may improve its functionality. Additional testing is required, and further work done with those developing the image-processing algorithms to ensure that the automatically generated annotations are of suitable

form and quality to be used in the system. Finally, this should be incorporated into a computer package which is usable on a desktop machine, so that it does become a useful tool for those attempting to read the Vindolanda texts.

6.2.1. Expanding Statistics from the Extant Corpus

The current system relies on data extracted from the Vindolanda ink tablet textual corpus, namely letter frequency information, a bi-gram analysis, a word list, and word list frequencies. There are other types of information that could be easily derived from this corpus and integrated into the system as it stands to possibly improve its functionality. For example, it would be simple to carry out a tri-gram, and quad-gram analysis, to investigate further the intrinsic statistical qualities of the letter sequences in the Vindolanda ink texts, but this information may not have any effect on system performance. Of more relevance would be the generation of common word endings from the existing corpus, and the incorporation of this information into the existing system. Additionally, expanding the available word list from the Vindolanda ink text corpus would provide more information for the system to utilize, and increase the likelihood of a word match being found.

6.2.2. Expanding the Word List

No attempt has been made to modify the word list derived from the ink tablets corpus (see S. 3.9): the word list of the system is currently limited to *exactly* what is featured in the Vindolanda textual corpus. This word list is presented as a 'left-to-right' or 'finite state' model of Vulgar Latin at Vindolanda, a type of language model which has been repeatedly shown to be deficient and inadequate since the early phases of psycholinguistic research (Chomsky 1957). The inability to generate possible new words from the existing data is, at present, a major fault in the system as 'It is important to bear in mind that the creation of linguistic expressions that are novel but appropriate is the normal mode of language use' (Chomsky 1972: 100). It would be possible to utilize the existing corpus to generate other instances of

text that could be found on the documents, thus expanding the data set available. This could be done by:

- Splitting existing words into fragments, and presenting these as possible letter sequences. This would be simple to implement.
- Developing some understanding of the underlying grammatical models at play during the formation of words. By doing so, it would be possible to generate new instances of words, for example, by identifying the root of a word, and suffixing a different ending to present the word as used in a different tense. However, it would be a complex task to implement this computationally. It may be easier to generate a derived word list manually.

It should be easy to write a small routine which provides fragments of the existing words as possibilities to the system (many of the texts that are encountered only have a section of words extant, and the papyrologist often completes the word, as shown in S. 2.5.5.3). For example, the word USSIBUS may be split into many different fragments, such as USS, USSI, USSIB, USSIBU, SSI, SSIB, SSIBU, SSIBUS, etc. These additional word fragments may very well match text on the stylus tablets, when the whole word is not present due to damage. This is an important limitation of the system at present, as it will only identify words which are present in the word list (although, if a word is not present in the word list, the sequence of characters presented as a solution is the most likely, due to the bi-graph information held within the system, as demonstrated in S. 5.2). Due to the fragmentary nature of the Vindolanda stylus tablets there is a high probability that complete words from the word list, or the exact match of fragments already present, would not be found.

In order to mimic the papyrologists' use of language, it will be necessary to integrate into the system some simple word creation mechanism. This would rely on the existing word list, statistics, and grammatical representation, to provide the basis of new words. This is a daunting task, given that 'even after constant research by hundreds of great minds, linguistics still lacks an adequate representation of *English*! Linguists know even less about other tongues' (Stainton 1996: 135). However, due to the way Latin uses word endings, it should be possible to construct a rudimentary model of how these work, and apply them to root forms of words, to generate possible

new words. Although there has been some success recently in modelling language systems to create new words using grammatical rules (Plunkett 1995), this would be a complex project. It would also be hampered by the fact that the grammar contained within the Vindolanda texts is not that of standard Classical Latin, and that the word corpus itself is perhaps too fragmentary to generate any concrete grammatical rules regarding the use of language at Vindolanda. One possibility is to loosely utilize research which has been done on the syntax and grammar of Classical Latin (Mahoney 2000), although it would be debatable how applicable this would be to the words in the database. It is therefore a task that should not be undertaken lightly.

An alternative solution would be to generate a derived word list manually. The experts could be questioned about the word list, and the knowledge engineer could manually work through the corpus, constructing different forms and tenses of verbs, plural nouns, adjectives, etc. This could provide an alternative word list, which would provide additional material for the system (although there would not be any statistical information available regarding the occurrence of words in the corpus).

6.2.3. Further Linguistic Work

If the sample set had been larger and less fragmentary, it would have been possible to gain an understanding of the word order, and higher level grammatical models at play, within the language of the Vindolanda documents. This could have been done by grammatically labelling the corpus on a word-by-word level and studying the patterns that emerge from the text (Lawler and Dry 1998). This qualitative data could be used to predict what *type* of word was needed next in a document, and so could narrow down the search for an appropriate match. However, because of the fragmentary nature of the corpus, no work on grammar was undertaken. It is something that should be kept in mind for the future, but it is not a priority in this research.

A further possibility with regard to providing alternative data sets would be to gather together groups of words that are linked by subject, for example, all the words in the corpus with respect to the

Roman army; food; transport; horses; leather goods; proper names; place names; numerals; dates; medicinal goods; livestock; trading; slaves; etc. Lists of words by subject are a tool commonly used by papyrologists (André 1981, 1985, 1991), and there is no reason why this could not be integrated into a computer system to provide these in electronic format. Locating words which belong to a probable subject could narrow down the search for a possible word fit, as, when directed by a human expert who was able to predict the context of a word, the system could search within a subset that would be relevant to the unknown word. This may or may not be helpful to the experts—perhaps if they already know the context of the word they are searching for, they would be able to identify the unknown section themselves! The construction of such word sets would involve a great deal of research, and would have to be undertaken manually. An investigation into improving the contextual XML markup of the Vindolanda texts is currently under way at the Centre for the Study of Ancient Documents (Hippisley 2005), and this work may result in a more interactive XML edition of the texts which would enable the automatic output of word lists from the corpus which pertain to certain topics.

6.2.4. Including Character Information

The character models currently used within the system are based solely on the stroke data that was captured during the annotation process. There were many other types of information that were captured when annotating the corpus. For example, it was noted when strokes went above and below the normal writing line: ascenders and descenders being one of the key features that the experts rely on to identify individual characters.² It should be possible to integrate this information into the present system, which may or may not improve its functionality, but would be worth investigating. Additional character models could be constructed using the manually added textual codes, to build up a textual description of individual characters. These could be compared to the output of the feature

² This has also been shown to be the case when reading clearly printed text: the shape of words (whether they have ascenders or descenders, etc.) affects the time taken to recognize them (Rayner 1998).

detection agents in the same way as present, by generating XML files containing information regarding the characters. Additional character sets could also be developed to cope with the difficulties of different forms of letters, as described in section 5.6.

6.2.5. Testing and Evaluation

The system needs to be tested on more complex lines of text, to see how it copes with more complicated data. This is particularly the case with the stylus texts, as the example shown in section 5.4 was a fairly simple one. Additionally, the system needs to be tested utilizing output from the image-processing algorithms generated from images of the stylus tablets. Considerable work on the image-processing techniques is necessary to allow clear annotations to be generated automatically so that these can be used by the system. The research presented here has demonstrated a proof of the concept: the key to developing a useful and usable computer package for experts lies in the fine tuning, testing, and development of the system.

6.3. APPLICATION DEVELOPMENT

The primary purpose of undertaking this research was to aid in the construction of an application that would assist the papyrologists in their reading of the Vindolanda stylus tablets. Although this monograph has detailed the success so far of this project, there is still a significant amount of work to be done regarding image processing and integration with the GRAVA architecture before a stand-alone system can be delivered to the experts. However, when those obstacles are overcome, it will be simple to utilize the architecture developed in this monograph as the basis of such a system. Because the GRAVA system is written in Yolambda, a dialect of Lisp, it will be easy to deliver the application using Java to end users due to its object oriented and flexible nature. There are a number of features that could be built into the interface to assist the papyrologist, and these are detailed below.

The interface must allow the expert to interact with the system. This has never been an attempt to create a system which can automatically 'read' texts without any input from a human user, although the architecture presented can reason independently from the user. This interaction can be achieved in many ways.

- The system should provide a 'one stop shop' so that an expert can load in an image, have the image-processing algorithms detect candidate strokes, and then propagate possible solutions based on this data and the statistics held within the system. Interaction should be possible with both processes, to change parameters and highlight information during the image processing, and to guide the system through the interpretation of the annotations. Selection of the data sets will be possible (if more than one word list is available, for example). It is the expert who should be in control of the system, not the other way around, although care will have to be taken with the development of the interface to allow the integration of the different information sources in a clear and usable fashion.
- Strokes will be able to be traced manually, allowing the program to reinterpret the stroke data and propagate further solutions.
- The image-processing parameters will be adjustable in order to reannotate less clear areas in a different manner, allowing the system to propagate alternative solutions.
- Any suggestions which may seem unlikely to the user could be discarded. The system could then propagate further alternative solutions.
- The expert should be able to see as many possible solutions as needed, not merely the one which has been deemed the closest. This could provide clues with which to obtain a true reading of the text.
- Solutions could be suggested by the user, with the system calculating the probability of these matching the stroke data present.
- The system should keep track of all changes made and tools used. This will provide a record of the reasoning process the expert undertook: something that is missing from the current documentation. This information should be available in a clear format, and retained for future use.

- An 'undo' function should be incorporated into the system to return it to a prior state, all the while keeping track of the reasoning process undertaken.
- Images of the annotated text should be easy to output to retain a visual record of the reading.
- A recursive mechanism would be put in place, where once a character or word has been identified they are added to the data sets to increase the information available for future runs of the system.
- The user should be able to add words to the word list as s/he sees fit, to increase the amount of relevant information available. Records would be maintained of any information that is added in this manner.
- By default the system should be set to run using data from the original word list. Additional corpora will be kept separate to preserve the integrity of the data.
- The user interface should be intuitive. A comprehensive help menu and full documentation should be provided.
- The program should incorporate commonly used image-processing tools, such as zoom, inverse, contrast, and brightness control, so that the experts can view images in the manner that they are accustomed to. Although these tools are already available in PhotoShop, incorporating them into the package would end the need for switching between different programs.

Work is currently being undertaken in the Department of Engineering Science at the University of Oxford to attempt to bundle the existing system into a package which can be used by the experts. The recommendations made above will be integrated into the future system, with full user testing undertaking to develop a package, which, it is hoped, will aid the papyrologists in their task.

6.4. EVALUATION

In Chapter 1 (S. 1.5), a list of six points was presented regarding the work which would be undertaken to build a system to read in images of the Vindolanda texts, and output plausible interpretations of those

texts. It is acknowledged that the final step has not been achieved yet (making an application available to the papyrologists). However, as described above, development is now under way. This should not detract from the success of this research: the work described in this monograph indicates how fruitful utilizing and developing computational techniques and technologies can be to construct computer systems to aid scholars in the humanities. There remain some points to be made, from the engineering and computing viewpoint, on the success of this research.

The use of an MDL-based architecture has demonstrated that it is possible to build a large system that can reason about different types of complex data efficiently, propagating useful solutions to interpretation problems. In effect this has paved the way for the construction of a 'cognitive visual system': one that can read in image data, and output useful interpretations of that data. Although there remains work to be done to dovetail this system with the feature detection image-processing algorithms, the success of this research indicates that utilizing an MDL-based architecture in this manner provides the framework necessary to build complex knowledge based image processing systems.

The development of special purpose image analysis tools has assisted in the identification of stroke information necessary to support hypotheses about the presence of characters. While the image analysis by itself is inadequate for reading the texts, it has been demonstrated that, when the stroke knowledge is fused with other constraining linguistic knowledge, plausible readings of tablets can be generated.

Because the foundations upon which the architecture is built (MDL and Monte Carlo sampling) are well understood, the behaviour and performance of the architecture can be estimated by means such as convergence analysis. While this application has no time critical aspect, in other applications where time may be an issue, such as real-time vision applications, the number of samples can be adjusted to fit the computational budget. In essence, interpretation accuracy can be traded off against time.

The agents in the MDL architecture appear to cooperate, but this cooperation is an emergent property of the architecture. The agents are explicitly *competitive* but the effects of Monte Carlo sampling

makes their aggregate behaviour appear to be *cooperative*. This emergent cooperative behaviour is perhaps the most important aspect of the architecture because it provides an implicit information fusion model that depends upon the effect that local decisions have upon the global interpretation. In the case of reading the Vindolanda tablets, information from character strokes is fused with statistical language information including character frequency, character bi-graph frequency, and word frequency knowledge. Additional knowledge such as grammatical knowledge or even subject context knowledge could easily be added, incrementally, to the described system.

It can be objected that this approach is computationally expensive but upon closer examination the comparison with competing techniques is encouraging. Most AI architectures, as discussed in Chapter 5, search for the first solution that could be found in a depth first manner. The goal here is to find approximations to the *best* solution so comparison with those architectures is not appropriate. Speech recognition and statistical natural language processing often use variants of Hidden Markov Models to find the best solution. These methods are comparable and are faster than the Monte Carlo approach advocated in this monograph but they cannot easily be adapted to work with the disparate kinds of knowledge that GRAVA can handle. The extra price paid for the Monte Carlo implementation more than pays for itself by providing a flexible knowledge fusion model. Thus this research has provided a proof of concept: and shown that MDL is a viable technique to build architectures that reason across semantically different data sets.

This research has also contributed uniquely to Classics. First, by asking how experts operate, it has been possible to understand better the processes the papyrologists undertake when reading ancient texts, which may be helpful to other scholars who work on primary source documents. Second, through the use of knowledge elicitation techniques, this process, which had previously not been studied, has been made explicit, allowing a model to be proposed of how humans read and interpret very ambiguous texts, a topic of interest to psychologists. Additionally, the type of information that is used when identifying individual characters has been made explicit. This allowed a framework to be developed that can annotate written text (there is no reason why this cannot be used to annotate other forms of writing other than the

ORC contained within the Vindolanda texts). The corpus of ORC characters that were annotated during this research is the only electronic palaeographic resource of its kind regarding this form of handwriting, and so may prove to be of use to scholars. The corpus has already given some insights into the letter forms used at Vindolanda, by providing a means to compare the characters present on the ink and stylus tablets. The statistics derived from the Vindolanda ink text word corpus also provide a resource for scholars in the field.

Although this research did not deliver a stand-alone application for the papyrologists to use to aid in reading the stylus tablets, this was not the primary aim of the project. It has provided an understanding of the type of tools required by the experts, as well as implementing a system that can analyse image data and propagate useful interpretations. Further testing and development is necessary before a completed application can be made available, but the findings presented here provide the basis for the construction of such a system: a fruitful culmination of varied, interdisciplinary research.

6.5. THE WIDER SCOPE

The research presents many opportunities for future work, applied to the wider community. From a humanities angle, this type of computer tool could prove to be instrumental in reading various types of documents that were illegible to the human eye: the joining of image processing and linguistic information allowing many possible interpretations of data to be generated to aid experts in their task. It would not be difficult to adapt this system to other linguistic systems if the necessary statistics were available, and permission ascertained for digitization of the primary sources, or access given to existing digital images of artefacts. Just how useful such a tool actually is to those who read such primary sources remains to be seen, but papyrology as a field has so far embraced computer-based tools and resources rapidly. It is estimated that there would be great interest from the community.

MDL-based architectures could be used on any number of image-processing tasks, where complex information from other semantic levels needs to be used to interpret images. It has been shown here

that MDL provides the common currency to relate different types of information, and this could be investigated further. An architecture such as the one described in this monograph could be used to read other types of handwriting, but the architecture could be expanded much further, to incorporate other semantic levels, such as grammar and contextual information. MDL architectures could also be used for entirely different image interpretation problems, such as facial recognition, sign language interpretation, or medical image analysis, as long as procedural information from different semantic levels was available, obtainable, or computable, to allow complex hierarchical systems to be implemented. MDL architectures could provide the basis for the development of any type of computer-based interpretation system, for example: speech recognition (or production), the analysis and interpretation of physical processes (the monitoring of weather, water flow and direction, and the many indicative signs of volcanic or seismic activity), predicting the outcome in war gaming systems, etc. The scope for the appropriation and development of this type of architecture is almost limitless: the important point is that it provides a way of comparing and contrasting semantically different types of information fairly and efficiently to generate the best probable outcome from available data.

6.6. IN CONCLUSION

This research has shown that utilizing an MDL architecture can provide the necessary infrastructure as a basis to implement a system that can read and interpret images of the Vindolanda texts, and in particular, the stylus tablets. Various adjustments to the system have been suggested, such as integrating easily obtainable statistical information, and more complex consideration has been made as to how this system may function in the future. However, further development must be undertaken before an integrated desktop application has been developed which will be able to aid the papyrologists in their day-to-day tasks. This depends on future development of this system, and further research into the image-processing algorithms to allow integration with this process.

This monograph presents a novel approach to a complex problem, delivering a system that can generate plausible interpretations from images, in the same way that human experts appear to do, to aid them in their task. The monograph effectively draws together disparate research in image processing, ancient history, and artificial intelligence to demonstrate a complete signal to symbol system. In doing so, this research offers further opportunities to develop intelligent systems that can interpret image data effectively, to aid human beings in complex perceptual tasks.

APPENDIX A

Using a Stochastic MDL Architecture to Read the Vindolanda Texts: Image to Interpretation

Co-authored with Dr Paul Robertson

Dinner... opposite me, a very beautiful lady... Next to me, (it slowly dawns on me) a slightly crazy guy... He asks what I do, and I say I'm a writer. He tells me about good writers and bad writers of his experience. He laughs at bad writers. People who can't do cursive or who put a little extra loop on their fs or gs make him laugh. He knows some good and bad writers on television. Sam Donaldson, on TV, has good handwriting. 'How do you know?' asks the lady, who has not quite realised that the slightly crazy guy is a slightly crazy guy. He explains that you can just tell, when you see them on the television. The worst writer he ever saw was a wrestler on TV. You just knew that he would go around and around when he did his Os. Round and around and around.

(Neil Gaiman, *Online Journal* 2003)

This appendix explores and describes in technical detail the implementation of the artificial intelligence system which was developed to propagate readings of the Vindolanda texts. An agent-based system, GRAVA, was adopted as its basis, and here we discuss the background and architecture of the original system, exploring the concepts of minimum description length (MDL) and Monte Carlo methods, which make GRAVA so novel and attractive for our purpose. The objects that constitute the GRAVA system are detailed, and the Monte Carlo algorithm applied to agent selection is expounded.

The original system (discussed in Chapter 4) had to be developed to be able to cope with the more complex data presented in the Vindolanda tablets, and the development of more sophisticated agents is explained. Character models of stroke data from the annotated Vindolanda corpus

are calculated, to provide data the system can utilize to ‘reason’ about unknown input data, and the final system, which reads in annotated images of unknown text, and outputs realistic interpretations of those images, is presented. This appendix details the underlying computational structures that comprise the system: results from the system are available in Chapter 5.

A.1. The GRAVA System

In his thesis (2001) Dr Paul Robertson utilizes an agent-based system to read a handwritten phrase, implementing a multi-level hierarchical agent-based model. This is akin to the Interactive Activation and Competition Model for Word Perception as proposed by Rumelhart and McClelland (McClelland and Rumelhart 1981: 379; see also Rumelhart and McClelland 1982; McClelland and Rumelhart 1986*b*; and S. 2.8.2) where it is proposed that a word is read by identifying features, characters, and, finally, words in a recursive process where ambiguities are resolved. However, Robertson’s GRAVA (Grounded Reflective Adaptive Vision Architecture) system does not use Parallel Distributed Processing (PDP)¹ architecture (unlike the original implementation of Rumelhart and McClelland’s model). PDP remains difficult to implement because the approach, which aims to develop neural network architectures that implement useful processes such as associative memory, throws out traditional models of computing completely, requiring a new computational model based on neuronal systems (Rumelhart *et al.* 1986; McClelland and Rumelhart 1986*a*). An additional problem with PDP is that the behaviour of such a network is highly nonlinearly related to the parameter values; a small change in those values may lead to very different behaviour, and therefore results (although this may also be the case with MDL). In order to overcome this intrinsic difficulty it is generally

¹ Parallel distributed processing (PDP) is a robust computer architecture, thought analogous to the processing structures found in the human brain, which uses decentralized networks of simple processing units to construct neurally inspired information processing models. This decentralization spreads responsibility, and changes in the state of the system are accomplished by modifying the strength or the architecture of the connections between neural units. The approach became popular in the 1980s with the work of McClelland and Rumelhart, and the PDP Research Group (McClelland and Rumelhart 1981, 1986; Rumelhart and McClelland 1982, 1986). PDP is often used to model mental or behavioral phenomena, such as the Interactive Activation and Competition Model for Word Perception, which is used to model how the human brain identifies words quicker than random letters or letters in non-words (see S. 2.8.2). PDP is the prevailing form of computer architecture known as connectionism, and is also often implemented as a ‘neural network’ (see Ch. 1, n. 26).

considered that PDP requires a large training set of data to ‘fine tune’ the parameters used within the system; there are not enough data available to us regarding the letter forms of Vindolanda to be able to do this. Instead Robertson implements an architecture where the atomic elements are implemented as agents² (in Yolambda, a dialect of Lisp),³ using familiar programming practices, which retains a more conventional programming model than the PDP approach.

The GRAVA system,⁴ developed by Robertson, manipulates agents, and builds programs from them. The primary purpose of an agent in the GRAVA system ‘is to fit a model to its input and produce a description element that captures the model and any parameterisation of the model’ (Robertson 2001: 59). Agent cooperation can span semantic levels, allowing hierarchical stacking in the same way that is described in PDP. This enables the building of systems that exhibit semantic interaction, a well-understood hierarchical concept that allows the behaviour and performance of systems to be closely monitored and understood (using techniques such as convergence analysis).⁵ Such an architecture is well suited to interpretation problems, which can be solved by fitting an input to a series of models and computing the likelihood of these matching. Many interpretation problems have more than one possible solution, and by using such a system many solutions can be propagated; the best solution being the ultimate target.

Of most relevance to *this* monograph, Robertson shows how his architecture is an effective way of rendering hierarchical systems by demonstrating how his software can ‘read’ a handwritten phrase. The text in this case was a nursery rhyme (see S. 4.2) which required only a small training set to be correctly interpreted by the system. Given that the GRAVA system can be shown to be a robust and easily adaptable architecture which can work successfully with relatively small corpora, and given the limitations of the dataset available to train the system, it was chosen as the basis on which to develop a system to read the Vindolanda ink, and eventually, stylus tablets.

A.2. GRAVA System Architecture

Robertson’s original demonstration of the GRAVA system, which aimed to read a handwritten phrase, comprises of three levels of agents. The first-level agents detect low-level features of written text, such as end points and

² See Ch. 4 n. 1.

³ See Ch. 4 n. 2.

⁴ For a full specification of the GRAVA system architecture, see Robertson (2001), Appendix B.

⁵ See Ch. 6 n. 2.

junctions of characters. The second-level agent builds character models from the database of character forms, and compares these models with the evidence presented by the first-level feature detectors to output a possible character identification. The highest level agent finds evidence of words in the input by looking at strings of data generated from the character finding agent and comparing these to a given corpus of words. The corpus used is the complete nursery rhyme, from which statistical information is collected and from which models are constructed regarding characters and words.

There were four separate agents that combined to make the lowest level feature detection section of the model (Figure A.1). Each agent reported on features discovered within a character position:

- Top stroke end points (T, above). This agent reported on the number of stroke end points at the top of the character. For example, the letter 'N' has one stroke endpoint on the top and 'W' has two.
- Bottom stroke end points (B). This agent reports on the number of stroke end points at the bottom of the character. For example the letter 'A' has two end points at the bottom of the character and letter 'I' has one.

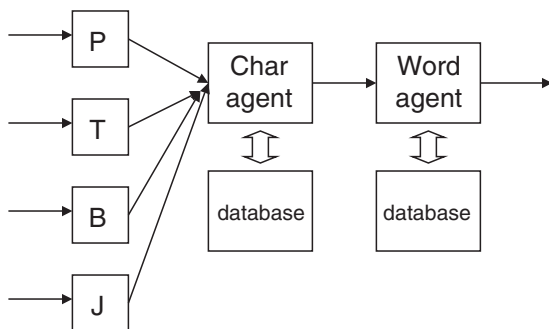


Figure A.1. Hand Written Phrase Recognizer Program. From Robertson (2001: 65). The 'P' agent detects that a character is present, i.e. it is not a blank space. The 'T' agent detects the presence of top end points, 'B' detects bottom end points, and 'J' detects junctions in the character. The combination of these features is then passed up to the character agent, who compares them against the combination of features of known characters in the database. A character is chosen and passed up to the word agent, which compares strings of these characters against known words in the database. A description of how well this string of characters matches the known words in the database is then outputted.

- Stroke Junctions (J). This agent reports on the number of line junctions formed from three or more lines. For example the letter 'A' has two such junctions. The letter 'N' has none.
- Character present (P). This agent detects whether the character position contains anything at all. Everything but the space character contains something (Robertson 2001: 64).

The character boxes are segmented into 'top' and 'bottom' so that the top stroke and bottom stroke features can be determined from a simple end point filter. Such a simple encoding system allows features to be flagged in order to pass the information up to the next character identification level: however, they are insufficient to unambiguously identify a character themselves. For example, the letters 'S', 'C', 'I', 'L', and 'N' all have one endpoint at the top, one at the bottom, and no junctions. In spite of the impoverished character representations used in this example, with the aid of semantics resulting from the appearance of characters in words the system is able to correctly identify the hand written text.⁶

The character level and word level both contain single agents to calculate the probability that the data in the corpus and the data presented match. GRAVA utilizes description length (DL) as a means of comparison (see Robertson 2001: 50, and S. 1.6.3 for the background of description length and minimum description length). Description length provides a fair basis for cost computation as it captures the notion of likelihood (or probability, 'P') directly: $DL = -\log_2(P)$. The minimum, or smallest, description length (MDL) generated when matching an input to a set of models yields the 'best fit' of that data to an individual model, and so presents the most likely match (as it is the most probable match of the unknown input to the model). The global description length generated by matching a series of inputs to models can be simply calculated by adding the DL of each unit, or

⁶ The idea of semantics affecting lower level processes is not new. In the late 1970s there was much interest within artificial intelligence concerning heterarchical-programming approaches (Fahlman 1973; McDermott and Sussman 1973). These systems suffered from behaviour that defied understanding and, equally importantly, analysis. Marr noted that in the human visual system that there was no evidence that higher level processes had any control over lower level vision and that we could profitably study low-level vision without addressing such issues. Computer vision has largely followed that path for the past 20 years. The issue has not gone away, however, and the notion of how differing levels of semantic interpretation are made to interact is at the very heart of the AI problem. The PDP approach is an interesting attempt to develop neural network architectures that implement useful processes such as associative memory. In this research, we aim for the higher level semantic information to override lower level noise in order to generate plausible interpretations.

output from each agent in the system. This can give a measure of the overall probability of the sequence, allowing easy comparison with other possible interpretations (the interpretation with the lowest global minimum description length being the most likely). MDL is used in GRAVA as a unifying method of comparing data from different levels of abstraction: thus providing a solution to the problem of how to compare the probabilities of both image and linguistic data matching the input.

Rather than present Robertson's results in this simple domain (which is discussed in S. 4.2) we will develop here our extension that supports interpretation of Vindolanda texts. First, however, we will describe in some detail description length and Monte Carlo methods, before discussing relevant parts of the GRAVA architecture.

A.2.1. Description length

The goal of the GRAVA architecture is to find an interpretation of its input data. In our case the input is stroke information (either hand generated, or automatically generated, see SS. 3.6 and 4.3.1). The interpretation is a structural description that includes subdescriptions for the words and character components. An interpretation, then, is a description that consists of many parts, each of which has an associated probability.

Consider the case of an image containing two blobs. If the probability of the first blob being a boy is given by $P(\text{blob}_1 = \text{boy}) = P_{\text{boy}}$ and the probability of the second blob being a dog is given by $P(\text{blob}_2 = \text{dog}) = P_{\text{dog}}$, the probability that the image is one in which the first blob is a boy and the second blob is a dog is given by $P_{\text{boy}} * P_{\text{dog}}$. An image containing n blobs such that for any $i < n$ the interpretation of blob_i is given by $I_i(\text{blob}_i)$ and the probability of such an interpretation being correct is given by $P(I_i(\text{blob}_i))$ the goal is to maximize the probability (assuming conditional independence):

$$\arg \max_{I_1, \dots, I_n} \prod_{i=1}^n P(I_i(\text{blob}_i))$$

Communicating the description through a noiseless channel using a theoretically optimal code would require a description length of: $(-\log_2(P_{\text{boy}})) + (-\log_2(P_{\text{dog}}))$. For an image containing n blobs as defined above the goal would be to choose interpretations $I_1 \dots I_n$ so as to minimize the description length:

$$\arg \min_{l_{1,n}} \sum_{i=1}^n -\log_2 (P(I_i(\text{blob}_i)))$$

Finding the most probable interpretation is identical to finding the minimum description length (MDL).

This approach raises two significant issues.

1. How to estimate P
2. How to find the global MDL.

Neither of these issues is straightforward. For the Vindolanda tablets we depend upon estimating P from frequencies obtained from a small corpus. For a general mechanism for finding the global MDL we employ a Monte Carlo sampling scheme discussed below.

A.2.2. Monte Carlo methods

The MDL agent architecture described in this appendix addresses the need to integrate knowledge at different semantic levels. To understand an image that consists of high-level components such as words, intermediate level features such as characters, and of low-level features such as strokes and end points, we need to integrate different levels of processing.

Single thread of control solutions, such as the blackboard⁷ and forward chaining⁸ approaches, depend upon taking a path towards a solution and backtracking past failures until a solution is found. The same is true of subsumption.⁹ These are essentially depth first searches for a solution—not

⁷ A blackboard system is an architecture for building problem-solving systems, in which a series of separate processes (or knowledge sources that generate independent solutions, such as expert systems) communicate through a centralized global memory or database, known as the blackboard. Partial solutions are built up and stored on the blackboard, which effectively coordinates the problem-solving process. (Illingworth 1996: 45). For example, the HEARSAY II blackboard system was developed for speech recognition (Erman, *et al.* 1980).

⁸ Also known as data-driven processing, forward chaining is a control procedure used in problem-solving and rule-based systems. ‘When a data item matches the antecedent part of a rule, then the conditions for that rule have been satisfied and the consequent part of the rule is executed. The process then repeats, usually through some form of conflict-resolution mechanism to prevent the same rules firing continuously’ (Illingworth 1996: 199).

⁹ Subsumption is a layered architecture predicated on the synergy between sensation and actuation, giving rise to all the power of intelligence, in lower animals such as insects. Brooks (1986) argues that instead of building world models and reasoning using explicit symbolic representational knowledge in simple worlds, we should follow the evolutionary path and start building simple agents in the real, complex, and

searches for the *best* solution. In a robot navigation problem, we may be happy if the robot negotiates the obstacles in the environment and finally ends up at the destination. In interpretation problems, just finding *a* solution is not good enough. A Latin sentence can have many plausible parses. Most of them do not make sense.

Markov Chain Monte Carlo Methods (MCMC)¹⁰ have become popular recently in computer vision (Geman and Geman 1984; Hammersley and Handscomb 1964; Lau and Ho 1997) but have been limited to modelling low-level phenomenon such as textures (Karssemeijer 1992). In natural language understanding, use of Hidden Markov Models (HMM)¹¹ has been successful as an optimization technique with certain restricted kinds of grammar. Problems that can be described as Hidden Markov Models (Baum 1972) can yield efficient algorithms. For example, in natural language understanding, some grammars permit efficient algorithms for finding the most probable parse. Stochastic Context Free Grammars (SCFGs)¹² can be parsed so as to find the most probable parse in cubic time using the Viterbi Algorithm¹³ (Viterbi 1967). Only the simplest grammars and problems can be solved efficiently in this way, however, and for the more interesting grammars and for more complex problems in general, other techniques

unpredictable world. No explicit knowledge representation of the world is used: the system is purely reflexive to input(s). Complex actions 'subsume' simple behaviours.

¹⁰ Markov Chain Monte Carlo (MCMC) is a computational technique long used in statistical physics, often used as a way to calculate Bayesian inference (an approach to statistics in which all forms of uncertainty are expressed in terms of probability). Probability is a well understood method of representing uncertain knowledge and reasoning to uncertain conclusions, so using MCMC provides a rich array of techniques that can be applied to solving problems in artificial intelligence which rely on computation regarding probability. See Neal (1993) for an introduction.

¹¹ A 'Markov Model' is a statistical model or simulation based on Markov chains, which are a sequence of discrete random values whose probabilities depend on the value of their predecessor. In a regular Markov Model, the state is directly visible to the observer, and therefore the state transition probabilities are the only parameters. In a Hidden Markov Model (HMM), the system being modelled is assumed to be a Markov process with unknown parameters, and the challenge is to determine the hidden parameters, from the observable parameters: hence the name Hidden Markov Model. See Rabiner (1989) for an introduction.

¹² A 'context-free grammar' is a formal system that describes a language by specifying a collection of rules that can be used to define how units can build into larger and larger structures. Stochastic context-free grammars are a version of context-free grammars with probabilities on rules, giving probabilistic descriptions of primary and secondary structures.

¹³ The Viterbi Algorithm (Viterbi 1967) is a dynamic programming algorithm used to find the most probable sequence of hidden states in a Hidden Markov Model, given a sequence of observed outputs.

must be used. Certainly something as loosely defined as an agent system incorporating semantics from multiple levels would rarely fit into the HMM straitjacket.

Even for natural language processing, finding the best solution can be prohibitively expensive when the Viterbi Algorithm cannot be employed. In visual processing, with images whose complexity greatly exceeds that of sentences, and which are three-dimensional (as opposed to the linear arrangement of words in a sentence), finding the best solution is infeasible. Approximate methods are therefore necessary: Monte Carlo¹⁴ techniques are very attractive in these situations.

In an ambiguous situation, such as parsing a sentence, in which many (perhaps thousands) of legal parses exist; the problem is to find the parse that is the most probable. If the problem can be defined in such a way that parses are produced at random, and the probability of producing a given parse is proportional to the probability that the parse would result from a correct interpretation, the problem of finding the most probable parse can be solved by sampling. If P is a random variable for a parse, the probability distribution function (PDF) for P can be estimated by sampling many parses drawn at random. If sufficiently many samples are taken, the most probable parse emerges as the parse that occurs the most frequently. Monte Carlo techniques use sampling to estimate PDFs.

Monte Carlo methods are attractive for a number of reasons:

1. They provide an approximate solution to the search of combinatorially prohibitive search spaces.
2. By adjusting the number of samples, the solution can be computed to an arbitrary accuracy.
3. Whereas the best solution cannot be guaranteed by sampling methods, measuring standard error during the sampling phase allows the number of samples to be adjusted to yield a desired level of accuracy.

GRAVA employs a Monte Carlo method: a means of providing approximate solutions by performing statistical sampling, to randomly choose which data is passed upwards between levels. This is a 'weighted random' selection process which picks likely data much more frequently than less likely examples (see Robertson 2001: 48). If only the data with the lowest description length were passed up between levels, the correct answer may never be found: the data with the locally minimum description length may not be the correct selection on the global level. The stochastic nature of this method of sampling ensures the generation of different results, and also means that the

¹⁴ See Ch. 1 n. 36 for a definition of Monte Carlo techniques.

system rapidly generates possible solutions without relying on exhaustive search (cutting search time). The system generates possible solutions on each iteration; the more iterations, the better the chance that a match to the solution is generated. Convergence on an ideal solution is then asymptotic:¹⁵ the system finds approximate solutions, and the more iterations that occur, the better the approximation that is reached. In practice, the system tends to find the exact solution in a short number of iterations, meaning that performance times are acceptable (this is demonstrated in Chapter 5).

A.3. Objects in the GRAVA architecture

The GRAVA architecture is built from a small number of objects: reflective layers; descriptions; description elements; interpreters; models; and agents. All of these terms are commonly used in cognitive science and artificial intelligence literature to mean a wide range of things. In the GRAVA architecture they have very specific meanings. Below, we describe what these objects are, the purpose, protocol, and implementation of the GRAVA objects, and how they cooperate to solve an interpretation problem, before demonstrating how the Monte Carlo algorithm chooses one element from a candidate list.

A.3.1 Reflective layers

A reflective layer takes an input description Δ_{in} and produces an output description Δ_{out} as its result. The output description is an interpretation ($I \in Q(\Delta_{in})$) of the input where $Q(x)$ is the set of all possible interpretations of x .

$$\Delta_{out} = I(\Delta_{in})$$

The goal of a layer is to find the best interpretation I_{best} that is defined as the interpretation that minimizes the global description length.

$$\arg \min_{I_{best}} DL(I_{best}(\Delta_{in}))$$

The interpretation function of the layer consists of an interpretation driver and a collection of connected agents. The interpretation driver deals with the formatting peculiarities of the input description (the input description may be an array of pixels or a symbolic description). The program is made from a collection of agents wired together. The program defines how the input will be interpreted. The job of the interpreter/program is to find the most

¹⁵ See section 5.3 for an example of the asymptotic nature of the GRAVA system.

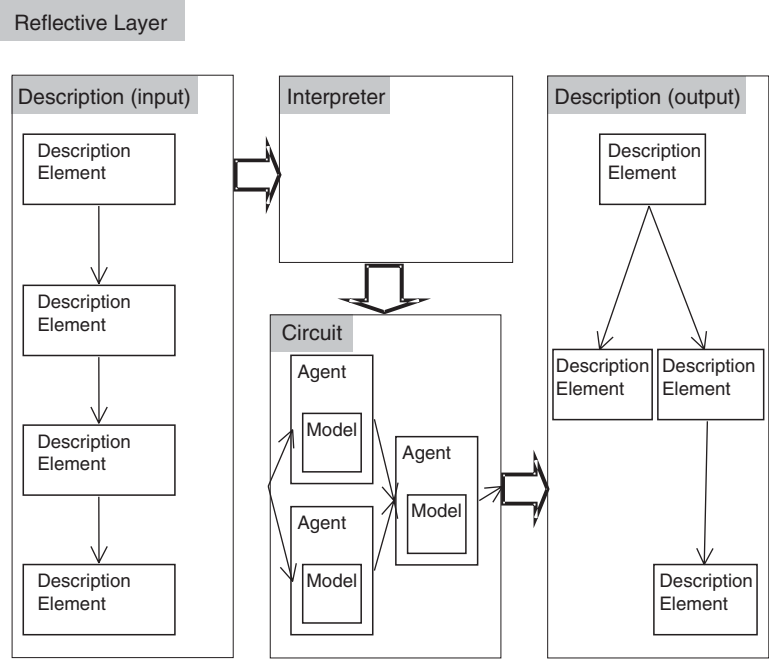


Figure A.2. Objects in the GRAVA architecture, indicating the relationship to description (input), interpreter, circuit, and description (output) within the reflective layer.

probable interpretation of the input description and to produce an output description that represents that interpretation.

A.3.2. Descriptions

A description Δ consists of a set of description elements ϵ

$$\Delta = \langle \epsilon_1, \epsilon_2, \dots, \epsilon_n \rangle$$

Agents produce descriptions that consist of a number of descriptive elements. The descriptive elements provide access to the model, parameters, and the description length of the descriptive element. For example, a description element for a face might include a deformable face model and a list of parameters that deform the model face so that it fits the face in the image. A description element is a $\langle \text{model}, \text{parameters} \rangle$ pair.

The description length is computed before an element is attached to a description because the agent must compete on the basis of description length to have the descriptive element included. It makes sense therefore to cache the description length in the descriptive element.

The description class implements the iterator:

(for Description|des fcn)

This applies the function ‘fcn’ to every element of the structural description, and this enables the architecture to compute the global description length:

$$DL(\Delta_{out}) = \sum_{i=1}^n DL(\epsilon_i)$$

The global description length of a particular set of descriptors, represented by the set ‘des’, is the sum of the description lengths computed for each descriptor ‘de’ in the set ‘des’. To get a better fix on notation, this is implemented in Lisp as:

```
(define (globalDescriptionLength Description|des)
  (let ( (dl 0))
    (loop for de in des
      (set! dl (+ dl (descriptionLength de))))))
```

A.3.3. Description elements

Description elements are produced by *agents* that *fit models* to the input.

Description elements may be implemented in any way that is convenient or natural for the problem domain. However the following protocol must be implemented for the elements of the description:

(agent <Element>) Returns the agent that fitted the model to the input;
 (model <Element>) Returns the model object that the element represents;
 (parameters <Element>) Returns the parameter block that parameterizes the model; and
 (descriptionLength <Element>) Returns the description length in bits of the description element.

Implementations of description elements must inherit the class DescriptionElement and implement the methods ‘agent’, ‘model’, ‘parameters’, and ‘descriptionLength’.

For readability we print description elements as a list:

(<model name> . <parameter list>)

A.3.4. Interpreters

An *interpreter* is a *program* that applies *agents* in order to produce a structural description output from a structural description input.

A.3.5. Models

Fitting a model to the input can involve a direct match but usually involves a set of parameters.

Consider as input, the string:

“three blind mice”

We can interpret the string as words. In order to do so, the interpreter must apply word models to the input in order to produce the description. If we have word models for ‘three’, ‘blind’, and ‘mice’ the interpreter can use those models to produce the output description:

((three) (blind) (mice))

The models are parameterless in this example. Alternatively we could have had a model called ‘word’ that is parameterized by the word in question:

((word three) (word blind) (word mice))

In the first case there is one model for each word. In the case of ‘three’ there is an agent that contains code that looks for ‘t’, ‘h’, ‘r’, ‘e’, and ‘e’ and returns the description element ‘(three)’. In the second case there is one model for words that is parameterized by the actual word. The agent may have a database of words and try to match the input to words in its database.

Consider the two examples above. If the probability of finding a word is 0.9 and the probability of the word being ‘three’ is 0.001, the code length of ‘(word three)’ is given by:

$$\begin{aligned} \text{DL}(\text{word three}) &= \text{DL}(\text{word}) + \text{DL}(\text{three}) \\ &= -\log_2(p(\text{word})) - \log_2(p(\text{three})) \\ &= -\log_2(0.9) - \log_2(0.001) = 0.1520 + 9.9658 = 10.1178 \text{ bits} \end{aligned}$$

The second approach, in which a separate agent identifies individual words would produce a description like ‘(three)’. The model is ‘three’ and there are no parameters. The likelihood of ‘three’ occurring is 0.001 so the description length is given by:

$$\text{DL}(\text{three}) = -\log_2(p(\text{three})) = -\log_2(0.9 \cdot 0.001) = 10.1178 \text{ bits}$$

That is, the choice of parameterized versus unparameterized does not affect the description length. Description lengths are governed by the probabilities of the problem domain. This allows description lengths produced by different agents to be compared as long as they make good estimates of description length.

A.3.6. *Agents*

The primary purpose of an agent is to fit a model to its input and produce a description element that captures the model and any parameterization of the model.

An agent is a computational unit that has the following properties:

1. It contains code, which is the implementation of an algorithm that fits its model to the input in order to produce its output description.
2. It contains one or more models (explicitly or implicitly) that it attempts to fit to the input.
3. It contains support for a variety of services required of agents such as the ability to estimate description lengths for the descriptions that it produces.

An agent is implemented in GRAVA as the class 'Agent'. New agents are defined by subclassing 'Agent'. Runtime agents are instances of the appropriate agent class. Generally agents are instantiated with one or more models. In our system all models are learnt from the corpus.

The protocol for agents includes the method 'fit' that invokes the agent's model fitting algorithm to attempt to fit one or more of its models to the current data.

(fit anAgent data)

The 'fit' method returns a (possibly null) list of description elements that the agent has managed to fit to the data. The interpreter may apply many agents to the same data. The list of possible model fits from all applicable agents is concatenated to produce the candidate list from which a Monte Carlo selection is performed.

A.3.6.1 Monte Carlo agent selection

A recurring issue in multi-agent systems is the basis for cooperation among the agents. Some systems assume benevolent agents where an agent will always help if it can. Some systems implement selfish agents that only help if there is something in it for them. In some cases the selfish cooperation is quantified with a pseudo market system.

Our approach to agent cooperation involves having agents compete to assert their interpretation. If one agent produces a description that allows another agent to further reduce the description length so that the global description length is minimized, the agents *appear* to have cooperated. Locally, the agents compete to reduce the description length of the image description. The algorithm used to resolve agent conflicts guarantees convergence towards a global MDL thus ensuring that agent cooperation ‘emerges’ from agent competition. The MDL approach guarantees convergence towards the most probable interpretation.

When all applicable agents have been applied to the input data the resulting lists of candidate description elements is concatenated to produce the candidate list.

The *monteCarloSelect* method chooses one description element at random from the candidate list. The random selection is weighted to the probability of the description element.

$$P_{\text{elem}} = 2^{-\text{DL}(\text{elem})}$$

So, for example, if among the candidates, one has a description length of one bit and one has a description length of two bits, the probabilities of those description lengths is 0.5 and 0.25 respectively. The *monteCarloSelect* method would select the one bit description twice as often as the two bit description.

The monteCarloSelect algorithm is given below:

```
(define (probability DescriptionElement|de)
  (expt 2.0 (- (descriptionLength de))))
(define (monteCarloSelect choices)
  (callWithCurrentContinuation
    (lambda (return)
      (let* ((sum (apply + (map probability choices)))
             (rsel (frandom sum)))
        (dolist (choice choices)
          (set! rsel (- rsel (probability choice)))
          (if (<= rsel 0.0) (return choice)))))))
```

A.4. Original Performance and the Need for Development

In the original implementation of his system, Robertson demonstrates how the low-level agents start out with a description of features based on the tops, bottoms, and junctions of characters. The system compares these with the models computed earlier from the corpus (on a character level),

computing for each symbol a description length. The character that is passed upwards to the next level is determined at random by the Monte Carlo Select algorithm. The system then compares the resulting ‘words’ with those in the corpus, generating description lengths for each. A global description length for the phrase is calculated by summing the description lengths of each of the symbols and words. In subsequent iterations different possibilities are generated, and the system searches for the best fit by looking for the configuration that gives the lowest global description length. This global minimum description length corresponds with a correct ‘reading’ of the text. A diagram demonstrating the system reading a handwritten text, *MARY HAS A LITTLE LAMB*, is shown in Figure 4.1, in Chapter 4.

In this way, Robertson shows that MDL formulation leads to the most probable interpretation, and also that global MDL mobilizes knowledge to address ambiguities. This demonstrates that, whilst the input data is inadequate for correctly identifying the characters unambiguously, the use of global MDL as a constraint allows correct identifications to be made. MDL also provides a ‘reasonable’ description in a relatively rapid time: exhaustive searches can fail to complete after several hours. The GRAVA system is therefore a robust, easily implemented, and adaptable architecture that produces convincing results in a short time frame, providing the ideal base for the development of a system to aid in the reading of the Vindolanda stylus tablets.

However, the original system required development because of the complexity of the data presented in the Vindolanda tablets, and the fact that eventually this system will dovetail with the image processing research done on the stylus tablets, reading in automatically annotated images. End point and junction data proved to be too unstable, given the noisy nature of the stylus tablets: small changes in image quality can create large differences in the number and location of the end points (this is demonstrated in S. 4.3.1). To counteract this, a more sophisticated character agent was developed which compares the stroke data of unknown characters to character models of known letters generated from the annotated corpus of images.

A.5. System Development and Architecture

The character database formed by annotating images of the ink and stylus tablets (as discussed in S. 3.6 and Terras and Robertson (2004)), is the main source of information regarding the character forms on the Vindolanda tablets. Models derived from this data can therefore be used to compare the unknown characters in the test set (which may be annotated by hand, or generated automatically, as described in S. 4.3.1). The difference between the

original system (SS. 4.2 and A.4) and the new system lies in the way that character models are developed, and how the test data is compared to this set of character models. Whereas the original system relied on end points, this final version relies on data regarding the strokes themselves. The end point agents in the feature level of the original system were replaced by a stroke detection agent. This results in models of characters that are less sensitive to the feature detection process (i.e. the generation of end points, which is problematic when dealing with noisy images such as those of the stylus tablets). It also means that the feature level agents depend on information that is much more easily propagated from automatic feature detection, allowing for easier amalgamation with the stroke detection system. Most importantly, stroke information is much more stable than end point data. Small changes in image quality cause only small changes in the stroke features detected. A schematic of the final system that was developed is shown in Figure A.3, incorporating all elements of the resulting process.

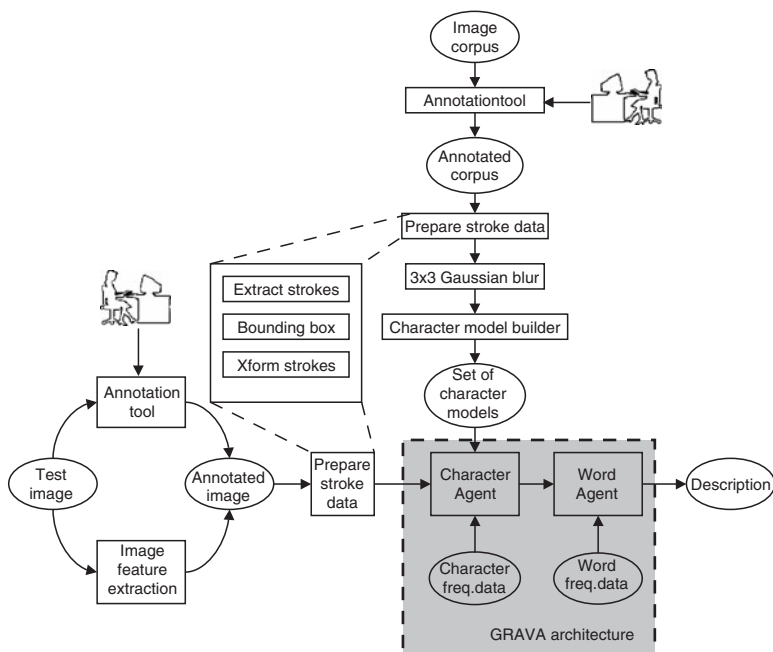


Figure A.3. Schematic of final system. Robertson's GRAVA architecture is highlighted to indicate the processes which are carried out as part of the final run of the system. The 'Annotated image' can be generated manually by an 'Annotation tool', or automatically by 'Image feature extraction'.

Before being used by the character agent, both the test set of data and the annotated corpus must be prepared in order to analyse their stroke data. This is done by extracting the strokes, drawing a bounding box around each of the characters to preserve groupings of strokes, and transforming these strokes onto a canonical sized grid to allow easy comparison.

Character models are built from the annotated set of images from the corpus by applying a Gaussian blur operator, and summing every instance of each individual character to build up a character model. The character agent then compares the unknown characters from the test data with those in the character models, also utilizing frequency information about each character to generate a description length for each model. One of these likely characters is then selected using the Monte Carlo sampling method and passed up to the word agent. This ensures that, over successive iterations, a fair, representative amount of each candidate character is selected and passed onto the next level.

The word level takes in the data from the character agent, combines them to form words, and compares them with the words from the corpus—the word ‘models’ in this case being the words found in the corpus. A description length for this comparison is noted. A selection is made from the possible words utilizing Monte Carlo sampling methods, and the final word output is generated.

The system then adds the description lengths for all the words in the phrase (or string of words) together, giving a global description length for that combination of characters and words.

The system repeats this process as often as the user dictates, and keeps track of the lowest global description length generated by each successive run. The smallest description length produced corresponds with the most likely answer: or the best fit answer available.

The preparation of the character models, and the way that both the character and word agents work, is discussed in detail below.

A.6. The Construction of Character Models

A character model is defined as a probability field that indicates the likely placing of one or more strokes of a two-dimensional character, producing a general representation of a character type. Unknown characters can then be compared to a series of these models, and the probability that they are an instance of each one calculated, the highest probability indicating a match. Whilst the first implementation of the system relied on an end point agent, this was replaced by a stroke detection agent that builds up character models based on the actual strokes of the character.

On a conceptual level, the (stroke-based) character model is constructed by taking an image of an individual character, finding its bounding box (identifying the rightmost x co-ordinate, and the leftmost x co-ordinate, and the highest and lowest y co-ordinates), and transforming this into a standardized (21 by 21 pixel) grid. The stroke data is convolved with a Gaussian Blur operator to reduce over-fitting. Each standardized representation is accumulated onto a generalized matrix for each character type: resulting in a generalized representation of each type of character. These are subsequently used as the models to which unknown characters are compared.

A.6.1. Finding the Bounding Box

The minimum and maximum x and y points of the character are noted, drawing a 'bounding box' around the letter. It does not matter how large or small this is, or how wide or narrow, as the representation is standardized by transforming this bounding box into a regimented size.

A.6.2. Calculating the Transform

The matrix presented by the bounding box is transformed into a standardized size to allow easy comparison of representations. This is achieved by simply translating the area within the bounding box to a canonical 21 by 21 pixel region at the origin.

$X_{\min}$, $Y_{\min}$ and $X_{\max}$, $Y_{\max}$ of the bounding box are noted. M_{width} and M_{height} are the model width and height respectively that the box is to be scaled to (in this case 21 by 21 pixels). A scaling matrix is then computed, giving the ratio to which the height and width will be scaled.

$$S_1 = \frac{M_{\text{width}}}{(X_{\max} - X_{\min})} \quad S_2 = \frac{M_{\text{height}}}{(Y_{\max} - Y_{\min})}$$

A translation vector is computed so that all characters are based at the origin. The scaling matrix is then applied to the translated stroke pixels, X_t and Y_t being the new scaled co-ordinates:

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} S_1 & 0 \\ 0 & S_2 \end{pmatrix} \left(\begin{pmatrix} X \\ Y \end{pmatrix} - \begin{pmatrix} X_{\min} \\ Y_{\min} \end{pmatrix} \right)$$

The choice of a 21 by 21 pixel grid was arrived at through a process of experimentation. If the grid is too large, then the data is too sparse and it is difficult to make any generalizations about the letter forms. The bigger the grid, the more examples are needed to make the generalized model applicable. The smaller the grid, the closer together the data, giving a trade-off

regarding size, accuracy, and usability. At first, an 11 by 11 array was used, but this was gradually increased, which improved the accuracy of this part of the system. The 21 by 21 sized grid seemed to give the best results for the training sets that were used.

A.6.3 Applying Gaussian Blur

The stroke data was convolved with a three by three Gaussian Blur¹⁶ operator. This creates intermediary points of data that reduces over-fitting of the model, to increase generalization from the examples given. Again, the three by three pixel field was chosen by experimentation.

A.6.4 Calculating the Final Model

The first character is drawn onto a ‘blank’ grid. A second instance of the character (if there is one available) is drawn over this, the values of this being added to the first instance. Additional instances of the character are laid over the grid, and the values summed as they go. This results in a composite model of all available character instances from the corpus, showing the path the strokes are most likely to make.

An example of how these steps combine to generate a character model is given in Figure A.4, where a small corpus which contains three ‘S’ characters is used to generate a character model of an S.

A.6.5. Learning Models from the Corpus

The corpus of annotated images (see S. 3.5) was randomly divided into a test set and two training sets: one ink tablet training set and one stylus tablet training set. Keeping the test and training data separate allows fair results to be generated. Two sets (ink and stylus) of character models were generated from the training set; these are shown in Figure A.5.

The character models generated from the ink data show that some of the letter forms, such as A, C, I, M, N, and T are fairly standardized throughout the test data. Other characters are more problematic. H, L, and V have a more confused appearance, indicating greater variability in the appearances of instances of the characters. D and Q are problematic, as the long strokes can either slope diagonally left to right, or right to left. The letter forms generated from this data can be compared to the letter forms generated from the stylus tablet data (see Appendix B and S. 3.11).

Not all of the letter forms are present in this selection, as the data set of the stylus tablets was much smaller than that of the ink tablets. However, some

¹⁶ A general purpose blur filter which removes fine image detail and noise leaving only larger scale changes.

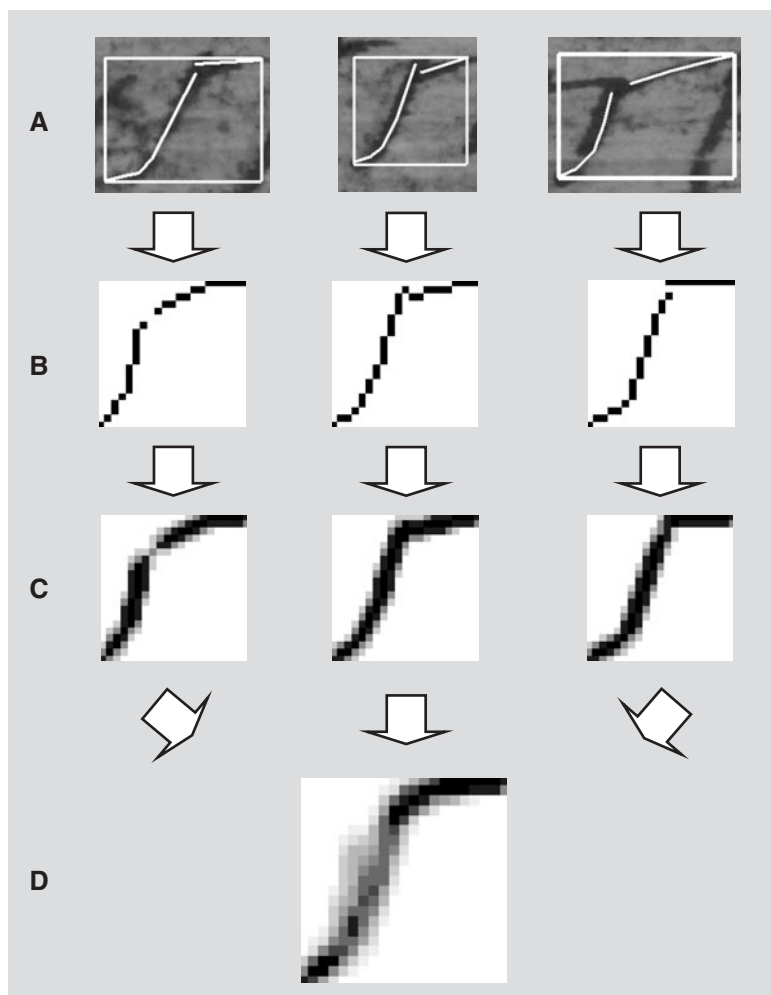


Figure A.4. The generation of a character model from a small corpus. Three letter 'S's are identified, and bounding boxes drawn around them (A). The stroke data is then transformed into a 21 by 21 pixel grid (B). A Gaussian Blur is applied (C). The composite images generate a character model (D). The darker the area, the higher the probability of the stroke passing through that pixel. In this way, the probabilities of the stroke data occurring are preserved implicitly in the character models.

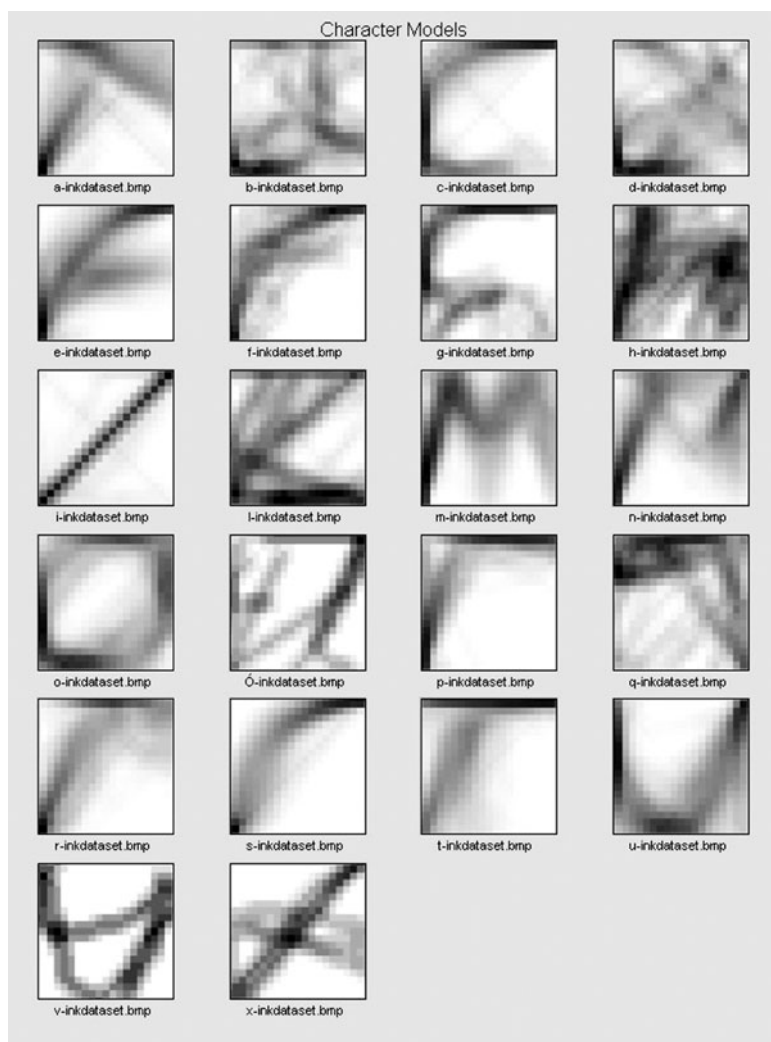


Figure A.5. Character models generated from the training set of the ink text corpus. The darker the area, the higher the certainty that a stroke will pass through that box. Shapes of the characters can clearly be seen, for example, the letter M.

interesting comparisons can be made between the two. The letter E shown is clearly a different form than that of the ink tablets (as discussed in S. 3.1.2). The letter I slopes in the reverse direction to that of the ink tablets (although

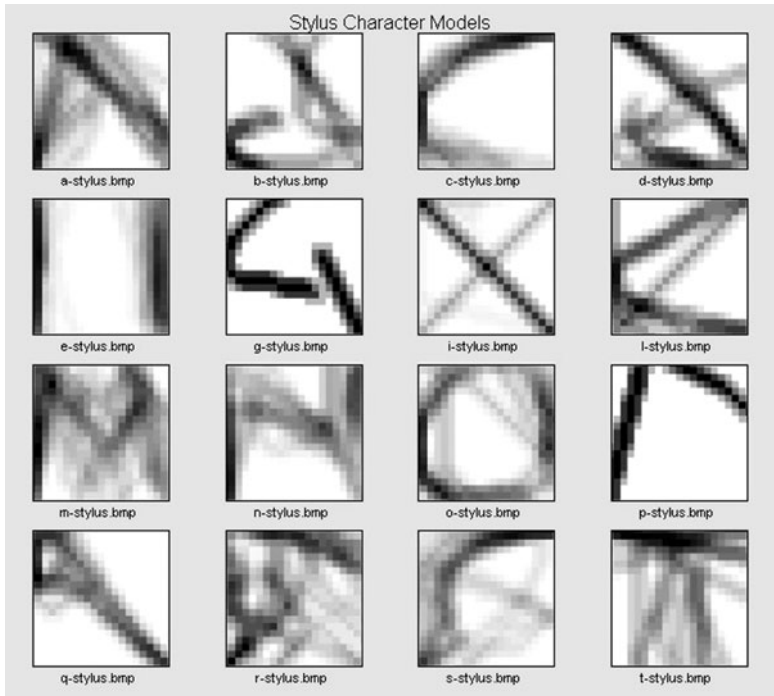


Figure A.6. Character models generated from the training set of the stylus text corpus. The darker the area, the higher the certainty that a stroke will pass through that box. These are more ‘clumsy’ compared to those generated from the ink texts: this is partly due to the small corpus used to generate these models. There is no letter F, H, U, V, or X in this training set of examples. The only example of the stylus letter U appears in the test data, rather than the training set (see 5.4 for how this problem is overcome within the system).

the forms of the characters are distorted due to the application of the transform, it still gives an indication of differences in the way characters are written). M, N, R, S, and T are much less standardized than in the ink tablets, perhaps because of the difficulties of writing in the different medium, and the small sample set available. Nevertheless, both sets of models provide adequate representations for comparison with unknown characters.

A.7. The Character Agent: Comparing Unknown Characters to the Character Models

The character agent’s function is to compare unknown characters to the character models composed from the training set. This is achieved by extracting

the strokes from the test data, transforming them to the standard size, then calculating the description length for matching the unknown character to each model in the data set. The character agent also utilises statistical information about the likelihood of a character being present, derived from the letter frequency analysis of the corpus. After the description length has been calculated (from a combination of the MDL frequency and MDL comparison of stroke evidence), one of the likely characters identified is selected, using the Monte Carlo sampling methods, and passed up to the word level.

A.7.1. Sizing and transform

Unknown characters are transformed to the standard size in the same way as described in sections A.6.1 and A.6.2: the stroke data is extracted, a bounding box is drawn around these strokes, and this matrix is transformed into a 21 by 21 pixel grid. No Gaussian Blur is applied to the stroke data in this case. This grid is a representation of the unknown character which can be compared with the standardised character model, calculated above.

A.7.2. Calculating the description length

Given the probability field representation of a character, we can calculate the probability that a line will cross through any given point. To generate possible identifications of the unknown character, we calculate the probability of how the evidence (stroke data) relates to each of the character models that have previously been constructed. Identification can never be certain, but it is possible to assign probability values, to calculate how the unknown data matches each character model.

We want to find the probability that the character is a , given the stroke evidence:

$$P(char = a|strokes)$$

The stroke data is assumed to be a legal character, so the sum of the probabilities for all the models should be 1:

$$\sum_{a \in \text{models}} P(char = a|strokes) = 1$$

However, $P(char = a|strokes)$ is not directly available to us. We have to calculate it utilizing Bayes' Theorem:

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)}$$

Therefore:

$$P(char = a | strokes) = \frac{P(char = a)P(strokes | char = a)}{P(strokes)}$$

$P(char = a)$, the probability of a character occurring, is already available to us from the lexicostatistics. $P(strokes | char)$ can be calculated from the model.

The model was made up by laying n different representations of a character over a standard grid, and adding up the occurrences of when the strokes passed through each box in the grid. To find the probability of one box in the grid being used as part of the character, simply take the value of the box in the composite model, and divide it by the number of characters which make the composite model.

$$P(Box_{xy} | char = a) = \frac{Model_a(x,y)}{nchars}$$

The stroke data can be viewed as the collection of boxes that the strokes pass through. Therefore, the total probability of the stroke data for each character is the product of the probability of the conjunctions of all the boxes passed through. If the stroke data goes through 'n' boxes, the probability of the stroke evidence given that the character is a is then:

$$P(stroke | char = a) = P(Box_1 \cap Box_2 \cap \dots \cap Box_n | char = a)$$

If we assume conditional independence,¹⁷

$$P(stroke | char = a) = P(Box_1 | char = a)P(Box_2 | char = a) \dots P(Box_n | char = a)$$

Which can be expressed as

$$P(stroke | char = a) = \prod_{i=1}^n P(Box_i | char = a)$$

¹⁷ This assumption is not entirely valid because the boxes are not strictly independent: there is some relationship between boxes as the strokes pass through them due to the trajectory of the pen or stylus stroke. However, the assumption works well in practice.

Giving

$$P(char = a|strokes) = \frac{P(char = a) \prod_{i=1}^n P(Box_i|char = a)}{P(strokes)}$$

$P(strokes)$ is a value such that

$$\sum_{a \in \text{models}} P(char = a|strokes) = 1$$

Therefore,

$$P(strokes) = \sum_{a \in \text{models}} \left\{ P(char = a) \prod_{i=1}^n P(Box_i|char = a) \right\}$$

This gives the final equation:

$$P(char = a|strokes) = \frac{P(char = a) \prod_{i=1}^n P(Box_i|char = a)}{\sum_{a \in \text{models}} \left\{ P(char = a) \prod_{i=1}^n P(Box_i|char = a) \right\}}$$

So the description length of the strokes using the model for 'a' is given by:

$$DL(strokes, a) = -\log_2 \left(\frac{P(char = a) \prod_{i=1}^n P(Box_i|char = a)}{\sum_{a \in \text{models}} \left\{ P(char = a) \prod_{i=1}^n P(Box_i|char = a) \right\}} \right)$$

The description length is computed for a comparison between all available character models and the unknown character. This is added to the description length of the probability of that character occurring in the sequence (generated from the lexicostatistics). Any character with a probability of less than 0.01 is discarded. This cuts down on processing time, and discards any truly unlikely character matches. The cut off value of 0.01 was arbitrarily chosen as an outlier rejection method. The Monte Carlo selection algorithm then chooses which of the remaining characters will be passed upwards to the next level, the description length of this character is, of course, passed up too.

A.8. The Word Agent: Comparing Unknown Words to the Word Corpus

The word agent's function is to compare strings of possible characters to the word models in the corpus, in much the same way as the character agent compares stroke data to models generated from the corpus. However, the word agent's task is considerably simpler than that of the character agent, as there is no need to represent stroke data. The word 'models' are the words in the corpus themselves (see S. 3.9), and the probability of them occurring is known from the word frequency data.

At the word level, word models consist of character strings (delimited by a space character at either end). The input passed up from the character level combines to make strings of characters, and these are compared to the word models on a letter by letter basis. If the letter matches that of the same position in the word model, it is flagged as a match. If the letter does not match, it is flagged as a non-match, and the description length from the character level is inserted in its place. The total description length of comparing that string of characters to the individual model is calculated as the negative logarithm (to the base 2) of the probability of the word occurring, plus the sum of the description lengths of the individual characters within the string (the DLs only being present if they did not match the associated character in the model directly):

$$DL = -\log_2 P(\text{Word}) + \sum_{i=1}^n \left\{ -\log_2 \left(\frac{2^{-DL(\text{strokes}, CH_i)} P(CH_{i-1} | CH_i)}{P(CH_{i-1})} \right) \right\}$$

if CH_i doesn't match, 0 otherwise

This is best understood by considering a possible message representing the fit of a word model to a character sequence. Consider fitting the word 'foo' to the word model 'for'. The message begins with a representation of the model for the word 'for', then for each character there is either a '1' bit, indicating that the character in that position matched the model, or a '0' bit indicating that the character didn't match, followed by a representation of the character that failed to match. For this example the message stream will be as shown in Figure A.7. The description length of the match is therefore the code length for the model used, one bit for each matching character, and one bit plus the code length for each non-matching character.

The system compares the string of characters to each individual word model in the corpus. The comparison with the lowest DL generates the most likely word identification.

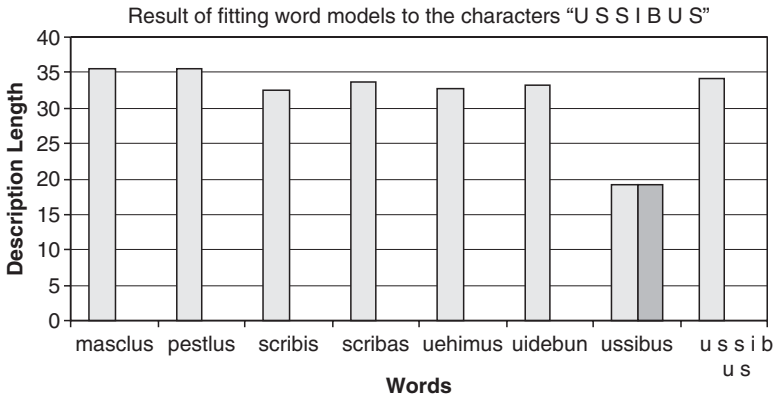


Figure A.7. Result of fitting word models to the characters USSIBUS, most likely fit. The word USSIBUS is by far the most likely contender compared to the other suggestions. The final ‘u s s i b u s s’ indicates the total description length if the description length for each individual character was added together: the overall MDL for ‘ussibus’ is smaller than that for ‘u, s, s, i, b, u, s’ indicating that the higher level interpretation at the word level is preferable to the locally correct interpretation at the character level.

However, what if the string of letters represents a word that is *not* in the corpus? The system compares the string to all available models, and also generates the total description length from the sum of all of the characters’ description lengths. If there is no match between the sequence and the existing models, this DL will be the lowest, and the solution is presented as a string of characters. However, due to the fact that bi-graph information is included in the word agent, the string with the minimum DL will be statistically the most likely combination of characters, even though a word has not been matched to the input. This can be seen in the results Chapter, 5.

A.8.1. An example

The string of characters ‘USSIBUS’ was fitted to all 342 of the word models that had a string length of seven characters. Only the results with a description length less than thirty-six bits are presented here (these being the lowest eight description lengths computed). It is clear that the string USSIBUS, present in the corpus, is the closest match, as this produces the lowest description length. Four other words (or word fragments present in the corpus) are identified as being as probable as the string of characters itself, SCRIBUS, SCRIBAS, VEHIMUS, UIDEUN. This indicates how the system propagates best-fit solutions, which are approximate to the correct solution.

Again, the system chooses which word to select as a solution using a Monte Carlo selection algorithm. The system does not incorporate any other data regarding word sequencing or grammar, at this stage. The MDL for a string of words is calculated by simply summing the description length for each word. Subsequent iterations of the system produce different sequences of characters, and therefore words. The most probable solution is that with the lowest MDL after a number of successive runs.

A.9. Conclusion

This appendix has detailed the technical underpinnings of the GRAVA system, and how it was developed to deal with the complex and more sophisticated data of the Vindolanda tablets. Explanations of minimum description length, Monte Carlo methods, and agent architectures have been provided, and the purpose, protocol, and implementation of individual objects in the GRAVA architecture have been discussed. The development of a more sophisticated character agent, utilizing character models ascertained from the annotated image corpus, is detailed, and complete overview of the developed system has been provided, with an example of how an individual agent works.

The MDL based GRAVA system provides a robust, adaptable platform from which to develop a system which solves an interpretation problem. Its success in generating plausible interpretations of the Vindolanda tablets in a realistic time frame is demonstrated in Chapter 5: there is no reason why this architecture cannot be developed and expanded to propagate possible solutions to other kinds of interpretation problems, and this is covered in Chapter 6.

APPENDIX B

Annotation

B.1. Encoding Scheme

The manual encoding scheme of additional tags is as follows. The tags are added, where necessary, to the ‘comments’ field of each annotated region.

These shortened tags were chosen for speed in annotation: it will be possible to expand these through simple textual transformation to make the comments noted for each individual stroke and character more human readable.

Direction (up, down, left, right) was judged from the central point of the individual character, on the base line. Assigning these directions was dependent on stroke movement away from this central point, as judged by the annotator.

Each *character* should have

- Letter identification (*)

- Overall letter size (S)

Each *stroke* should have (in this order, separated by a comma)

- Stroke direction (D)

- Stroke length (L)

- Stroke width (W)

- Place on line (P)

Each *stroke meeting* should have

- Angle (A)

These individual fields are expanded, below.

Letter identification (*)

- Letter (*a, *b, etc)

- If unidentified (*?)

- If expected, but not present, (*!)

Overall letter size (S)

Height (SH)

- Large (SHl)

- Average (SHa)

- Small (SHs)

Width (SW)

Large (SWl)

Average (SWa)

Small (SWs)

Direction of stroke (D)

Straight(DS)

Down Left (DSdl)

Down Right (DSdr)

Up Left (DSul)

Up Right (DSur)

Horizontal (DSh)

Vertical (DSv)

Curved (DC)

Simple Curve (DCS)

Down to Left (DCSdl)

Curve Left (DCSdlcl)

Curve Right (DCSdlcr)

Down to Right (DCSdr)

Curve Left (DCSdrcl)

Curve Right (DCSdr cr)

Up to Left (DCSul)

Curve Left (DCSulcl)

Curve Right (DCSulcr)

Up to Right (DCSur)

Curve Left (DCSurcl)

Curve Right (DCSurcr)

Double Curve (Wave) (DCW)

Down to Left (DCWdl)

Curve Left (DCWdlcl)

Up (DCWdlclu)

Down (DCWdlcld)

Curve Right (DCWdlcr)

Up (DCWdlcru)

Down (DCWdlcrd)

Down to Right (DCWdr)

Curve Left (DCWdrcl)

Up (DCWdrclu)

Down (DCWdrclld)
 Curve Right (DCWdrclr)
 Up (DCWdrclru)
 Down (DCWdrclrd)
 Up to Left (DCWul)
 Curve Left (DCWulcl)
 Up (DCWulclu)
 Down (DCWulclld)
 Curve Right (DCWulclr)
 Up (DCWulclru)
 Down (DCWulclrd)
 Up to Right (DCWur)
 Curve Left (DCWurcl)
 Up (DCWurclu)
 Down (DCWurclld)
 Curve Right (DCWurclr)
 Up (DCWurclru)
 Down (DCWurclru)
 Horizontal (DCWh)
 Curve Left (DCWhcl)
 Up (DCWhclu)
 Down (DCWhclld)
 Curve Right (DCWhclr)
 Up (DCWhclru)
 Down (DCWhclrd)
 Vertical (DCWv)
 Curve Left (DCWvcl)
 Up (DCWvclu)
 Down (DCWvclld)
 Curve Right (DCWvclr)
 Up (DCWvclru)
 Down (DCWvclrd)

 Loop (DL)
 To Left (DLl)
 Open (DLlo)
 Closed (DLlc)
 To Right (DLr)
 Open (DLro)
 Closed (DLrc)

Stroke Length (L)

Comparative

Short (Ls)

Average (La)

Long (Ll)

Stroke Width (W)

Comparative

Thin (Wt)

Average (Wa)

Wide (Ww)

Place on Line (P)

Within Line Average (Pw)

Descender (PD)

Below Left (PDL)

Below Right (PDr)

Ascender (PA)

Above Left (PAL)

Above Right (PAr)

Stroke Meeting Angle (A)

Open to Top (AT)

Obtuse (ATo)

Right (ATr)

Acute (ATa)

Open to Bottom (AB)

Obtuse (ABo)

Right (ABr)

Acute (ABa)

Open to Left (AL)

Obtuse (ALo)

Right (ALr)

Acute (ALa)

Open to Right (AR)

Obtuse (ARo)

Right (ARr)

Acute (ARa)
 Crossing (AC)
 Right Angle (ACr)
 Compressed Vertical (ACv)
 Compressed Horizontal (ACh)
 Perpendicular (AP)

B.2. File Format

Each annotation is preserved in XML file format. The annotation file for each image contains a single annotation tag `GTAnnotations`, which encapsulates all of the regions in the file, with the following attributes:

- *imageName*: The name of the source image file that this file annotates.
- *author*: The name of the last person to update the annotation file.
- *creationDate*: The date and time that the annotation file was initially created.
- *modificationDate*: The date and time that the annotation file was last modified.

Each individual region that is annotated is represented by a `GTRegion` tag which has the following attributes:

- *author*: The name of the person who created or last modified this region.
- *regionType*: An identifier (such as 'R26') that identifies the labelling of the region. The mapping of region types to human readable names as specified in the region dictionary file: these labels are explained below (S. A.3).
- *regionUID*: A unique region identifier (such as 'RGN93') that names the region.
- *regionDate*: The date and time that the region was created or last modified.
- *co-ordinates*: A list of pairs of numbers that represent the co-ordinates of points along the boundary of the regions. The boundary of the region is defined to be the region contained within the region formed by drawing straight lines between these points.
- *comments*: Additional tags manually added (S. A.1) to describe each stroke/region further.

(This text is an updated version of that found in Robertson 2001: Appendix C.1.)

An example of a full XML encoding of a character is given below. This text describes the letter S, as shown in Figure 3.16, first defining a character box (`ADCHAR0`), which gives the coordinates of the character box, which relate in pixels to the image specified in the `imageName` element. Additional

comments are added to this character box to specify that it is the letter S, and of a large height (*s, SHl). The two individual strokes are identified (SO1, SO2: this numbering is related to the order of strokes that the scribe was expected to make, although this cannot be certain) and each of these is given an abstracted description regarding their direction, length, and width in the comments field. Stroke ends are then identified: SE4 (Stroke ending with a hook up to the left), and SE1 (Stroke ending bluntly), and the junction is then marked (SMJ3: stroke meeting, cross meet where the strokes cross each other slightly at the ends) with an extra comment to indicate that the meeting is open to the right at an obtuse angle (ARo). A full table of all of these codes used is given in section B.3. Comments included in the file, below, are those which correspond to the list above (S. B.1).

```
<GTAnnotations imageName="C:\grava\311l.tif" author = "Melissa Terras" creationDate = "09/10/02 15:23:58" modificationDate = "09/10/02 15:27:30">
```

```
<GTRegion author = "Melissa Terras" regionType = "ADCHAR0" regionUID = "RGN0" regiondate = "09/10/02 15:24:36" coordinates = "338, 22, 298, 4, 186, 30, 106, 62, 94, 196, 36, 356, 46, 408, 140, 420, 184, 288, 198, 106, 282, 68, 338, 22" comments = "*s, SHl, SWl"></GTRegion>
```

```
<GTRegion author = "Melissa Terras" regionType = "SO1" regionUID = "RGN1" regiondate = "09/10/02 15:25:22" coordinates = "170, 76, 162, 184, 152, 276, 124, 350, 102, 382, 74, 372" comments = "DSdl, Ll, Wa, PDI"></GTRegion>
```

```
<GTRegion author = "Melissa Terras" regionType = "SO2" regionUID = "RGN2" regiondate = "09/10/02 15:26:11" coordinates = "146, 94, 286, 18" comments = "DSur, La, Wa, PAr"></GTRegion>
```

```
<GTRegion author = "Melissa Terras" regionType = "SE4" regionUID = "RGN3" regiondate = "09/10/02 15:26:38" coordinates = "62, 354, 94, 354, 94, 386, 62, 386, 62, 354"></GTRegion>
```

```
<GTRegion author = "Melissa Terras" regionType = "SE1" regionUID = "RGN4" regiondate = "09/10/02 15:26:44" coordinates = "162, 62, 178, 62, 178, 88, 162, 88, 162, 62"></GTRegion>
```

```
<GTRegion author = "Melissa Terras" regionType = "SE1" regionUID = "RGN5" regiondate = "09/10/02 15:26:50" coordinates = "134, 80, 154, 80, 154, 104, 134, 104, 134, 80"></GTRegion>
```

```
<GTRegion author = "Melissa Terras" regionType = "SE1" regionUID = "RGN6" regiondate = "09/10/02 15:26:56" coordinates = "270, 14, 294, 14, 294, 36, 270, 36, 270, 14"></GTRegion>
```

```
<GTRegion author = "Melissa Terras" regionType = "SMJ3" regionUID =
"RGN7" regiondate = "09/10/02 15:27:12" coordinates = "154, 70, 184, 70,
184, 98, 154, 98, 154, 70" comments = "ARo"></GTRegion>
</GTAnnotations>
```

B.3. Region Type Identifiers

Each region is given a region type identifier, to specify whether it is a type of character, stroke, end point, or junction. These identifiers are specified in Table B.1.

Table B.1. Region identifier codes.

Identifier	Definition
ADCHAR0	Character Box
ADCHAR1	Space Character
ADCHAR2	Paragraph Character
ADCHAR3	Interpunct
SO1	Stroke—First Stroke
SO2	Stroke—Second Stroke
SO3	Stroke—Third Stroke
SO4	Stroke—Fourth Stroke
SO5	Stroke—Fifth Stroke
SO6	Stroke—Sixth Stroke
SO7	Stroke—Seventh Stroke
SE1	Stroke End—Blunt
SE2	Stroke End—Hook—Down Left
SE3	Stroke End—Hook—Down Right
SE4	Stroke End—Hook—Up Left
SE5	Stroke End—Hook—Up Right
SE6	Stroke End—Ligature—To Left—Down
SE7	Stroke End—Ligature—To Left—Up
SE8	Stroke End—Ligature—To Right—Down
SE9	Stroke End—Ligature—To Right—Up
SE10	Stroke End—Serif—To Left—Down
SE11	Stroke End—Serif—To Left—Up
SE12	Stroke End—Serif—To Right—Down
SE13	Stroke End—Serif—To Right—Up
SMJ1	Stroke Meeting—Close Meet
SMJ2	Stroke Meeting—Exact Meet
SMJ3	Stroke Meeting—Cross Meet
SMJ4	Stroke Meeting—Midpoint—Close Meet
SMJ5	Stroke Meeting—Midpoint—Exact Meet
SMJ6	Stroke Meeting—Midpoint—Cross Meet
SMJ7	Stroke Meeting—Crossing

B.4. Viewing the Annotated Corpus

The annotated corpus can be easily viewed using a Java-enabled web browser (preferably Netscape Version 6.0 or above). These annotated images will eventually be presented on Vindolanda tablets online (hosted by the Centre for the Study of Ancient Documents, St Giles, Oxford, see <http://vindolanda.csad.ox.ac.uk/>). More work must be done on the user interface to make this a useful tool for papyrologists and palaeographers. Additionally, transformations will be applied to the XML files to expand the abbreviated comments into a more human readable form. The comments could then be used for further palaeographic analysis of the letter forms contained within the Vindolanda texts (see Terras and Robertson 2004: 409) for a demonstration of how a comparison of the comments can add to our understanding of the letter forms. This system has the potential to be a rich training and research set for palaeographers and papyrologists.

The user is presented with a list of all 110 annotated images (Figure B.1). The images of the documents are split into line by line sections, to allow for easier annotation. The last number in the name of the file indicates the line it represents in that document. By double clicking on one of these files, the annotated section becomes visible. (Figure B.2). It is possible to toggle the annotations on and off to compare the underlying images to the annotations. (Figure B.3).

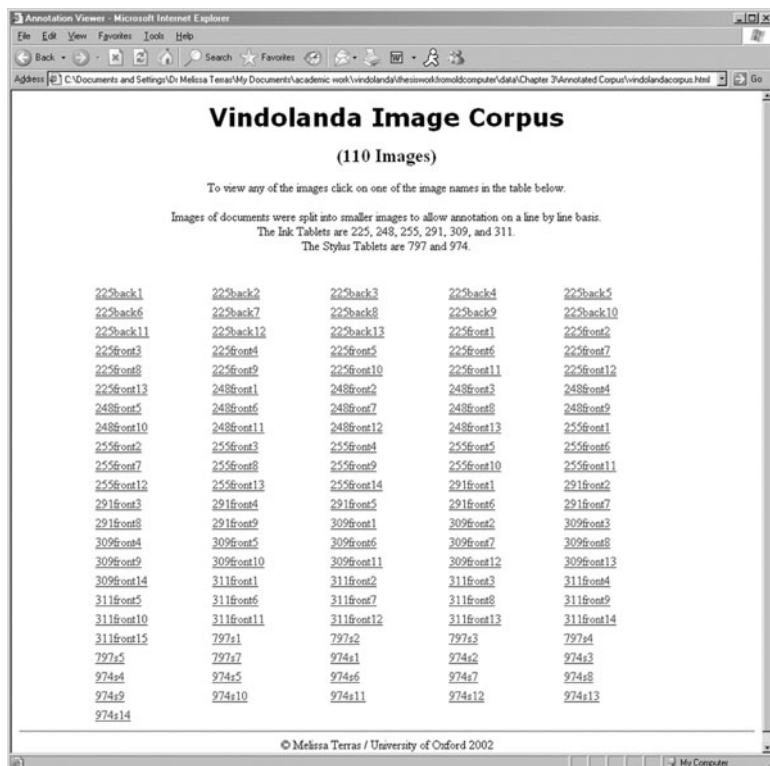


Figure B.1. Opening Vindolanda Image Corpus Screen, as seen through the browser.

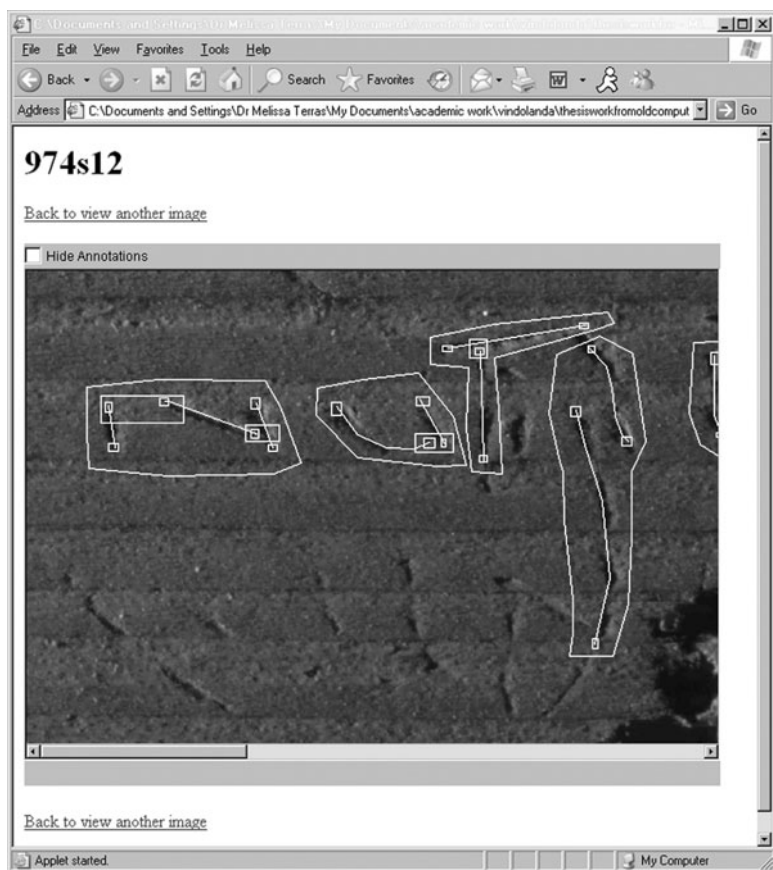


Figure B.2. Annotated section of stylus tablet 974, line 12, showing 'NUTR' and the related features highlighted. See Figure 2.4 for an image of the complete text.

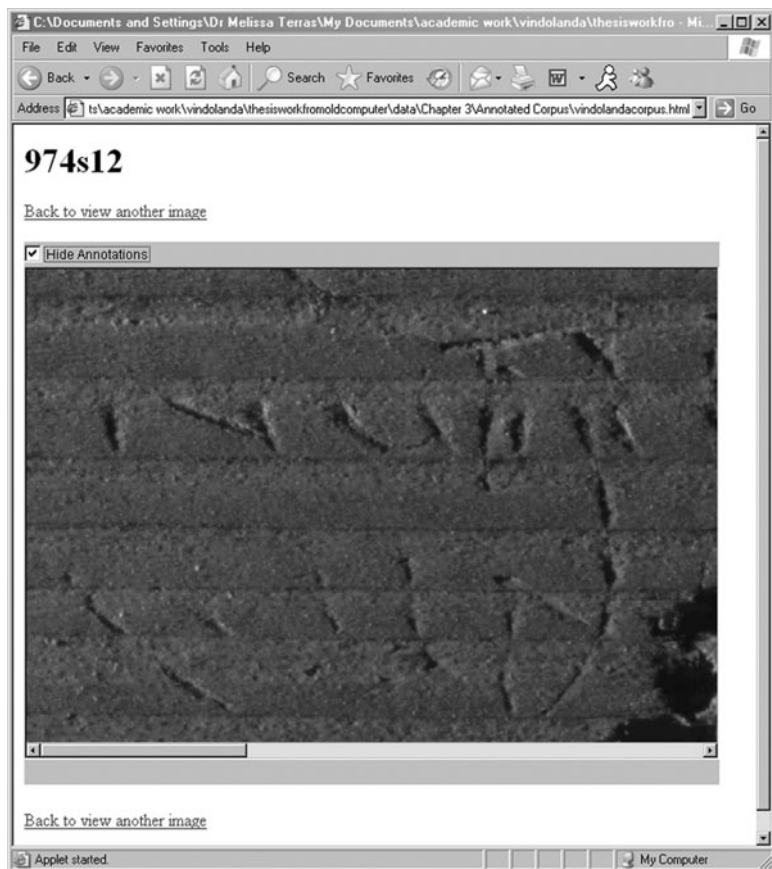


Figure B.3. Section of tablet shown above, with annotations toggled off. This allows the user to switch between their interpretation of the text, and that implied by the markup.

APPENDIX C

Vindolanda Letter Forms Corpus

This appendix contains a visual representation of the stroke data that was captured when annotating the images of the Vindolanda ink and stylus tablets. Each character is presented individually, first showing the standard letter form as identified by the papyrologists (Bowman and Thomas 1983), then the character models that were generated from the ink and stylus tablet data,¹ to allow comparison with the standardized forms. Finally, every instance of the characters in the data set is shown here,² to give a indication of the form of the characters that were found in the documents.³ Of course, the complete set of annotated images can be viewed using a web browser, as described in Appendix B.

There were no instances of F or H in the training corpus or the test corpus of annotated stylus tablets. There were no instances of U or V in the training corpus of annotated stylus tablets but some examples in the test corpus and these are shown. The standard representation of the character X was provided by one of the experts in the knowledge elicitation exercises, rather than from Bowman and Thomas 1983. There were no instances of X in the training corpus or the test corpus of annotated stylus tablets.

¹ These images were generated by Dr Paul Robertson.

² These images were prepared with the help of Dr Xiabo Pan, formerly of the Robots Research Laboratory, Department of Engineering Science, University of Oxford, now at Mirada Solutions Ltd.

³ The characters U and V are somewhat interchangeable, but are both present in the data set due to the fact that they were used by Bowman and Thomas in the Vindolanda texts, and each U or V character was annotated according to this text.

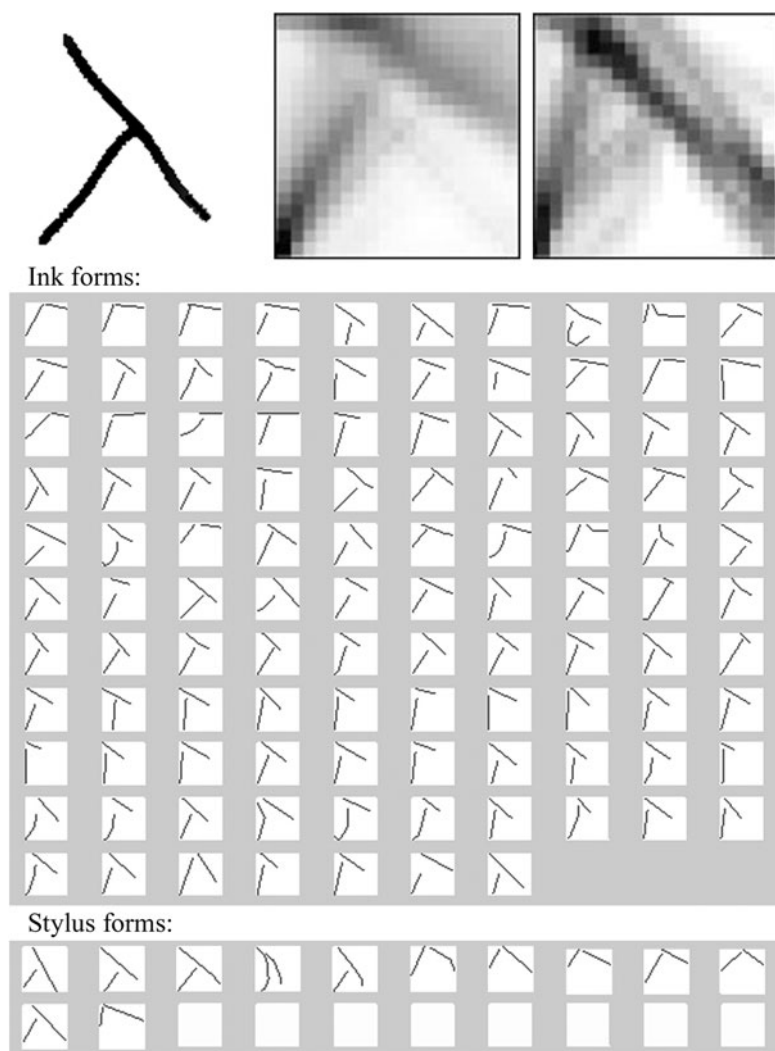


Figure C.1. The standard representation of the character A given in Bowman and Thomas (1983). The character model of the character A generated from all the ink characters in the training corpus, and the character model of the character A generated from all the stylus characters in the training corpus. Every instance of the character A in the ink tablet annotated corpus. Every instance of the character A in the stylus tablet annotated corpus.

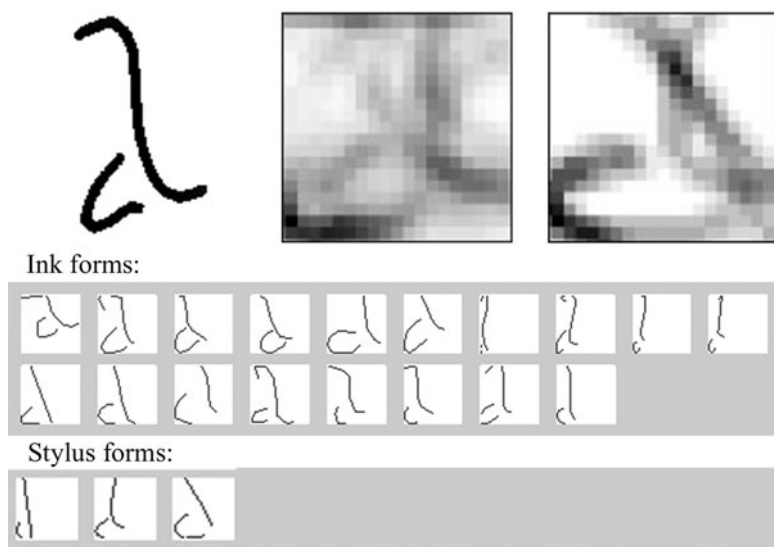


Figure C.2. The standard representation of the character B given in Bowman and Thomas (1983). The character model of the character B generated from all the ink characters in the training corpus, and the character model of the character B generated from all the stylus characters in the training corpus. Every instance of the character B in the ink tablet annotated corpus. Every instance of the character B in the stylus tablet annotated corpus.

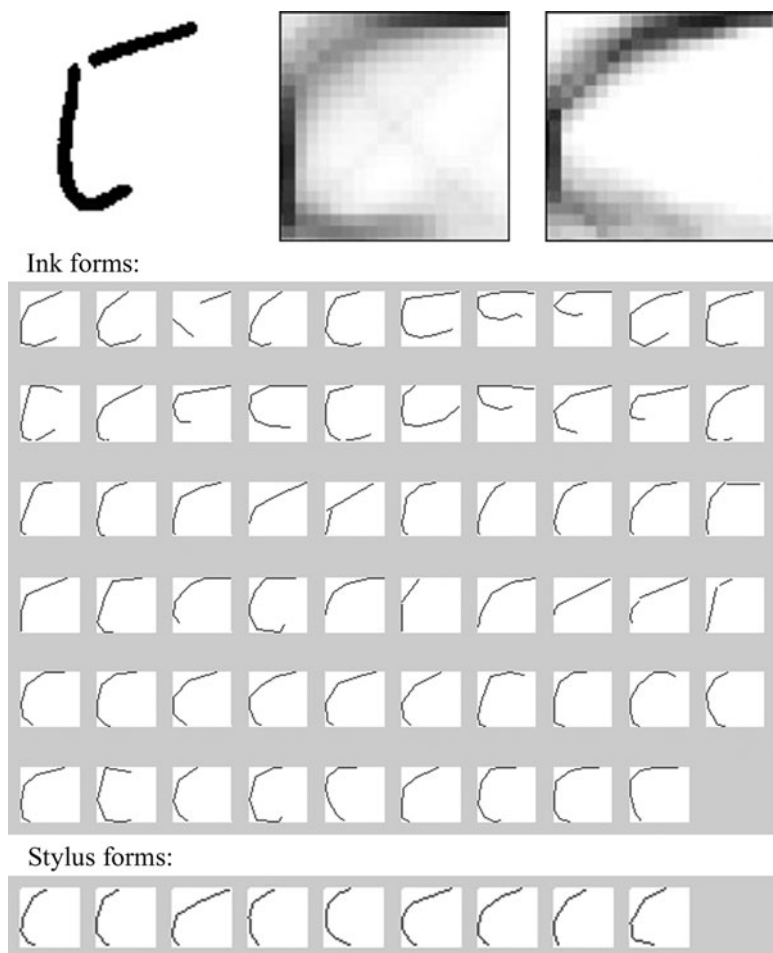


Figure C.3. The standard representation of the character C given in Bowman and Thomas (1983). The character model of the character C generated from all the ink characters in the training corpus, and the character model of the character C generated from all the stylus characters in the training corpus. Every instance of the character C in the ink tablet annotated corpus. Every instance of the character C in the stylus tablet annotated corpus.

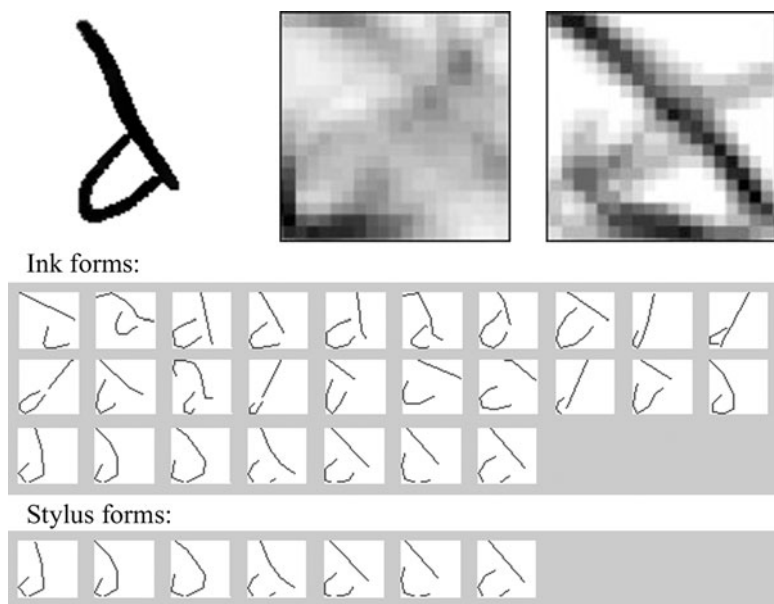
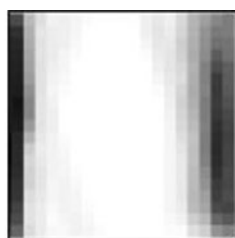
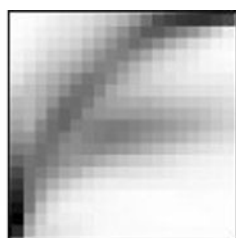
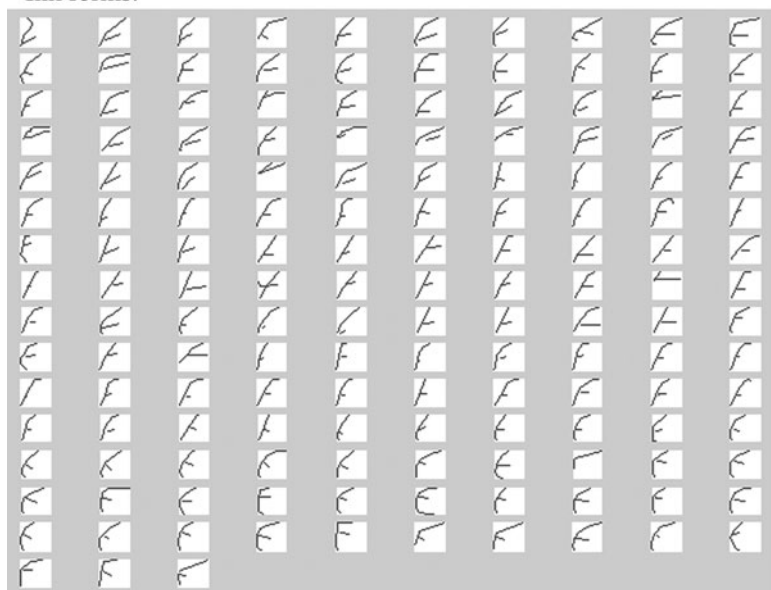


Figure C.4. The standard representation of the character D given in Bowman and Thomas (1983). The character model of the character D generated from all the ink characters in the training corpus, and the character model of the character D generated from all the stylus characters in the training corpus. Every instance of the character D in the ink tablet annotated corpus. Every instance of the character D in the stylus tablet annotated corpus.



Ink forms:



Stylus forms:



Figure C.5. The standard representation of the character E given in Bowman and Thomas (1983). The character model of the character E generated from all the ink characters in the training corpus, and the character model of the character E generated from all the stylus characters in the training corpus. Every instance of the character E in the ink tablet annotated corpus. Every instance of the character E in the stylus tablet annotated corpus.

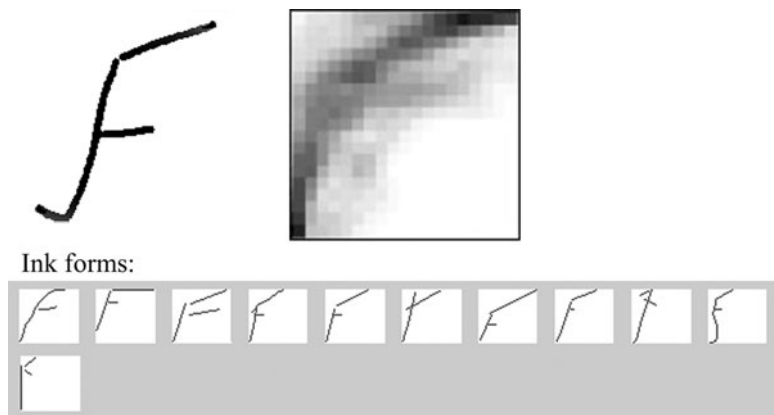


Figure C.6. The standard representation of the character F given in Bowman and Thomas (1983). The character model of the character F generated from all the ink characters in the training corpus. There were no instances of the character F in the training corpus, or the test corpus, of the annotated stylus tablets. Every instance of the character F in the ink tablet annotated corpus.

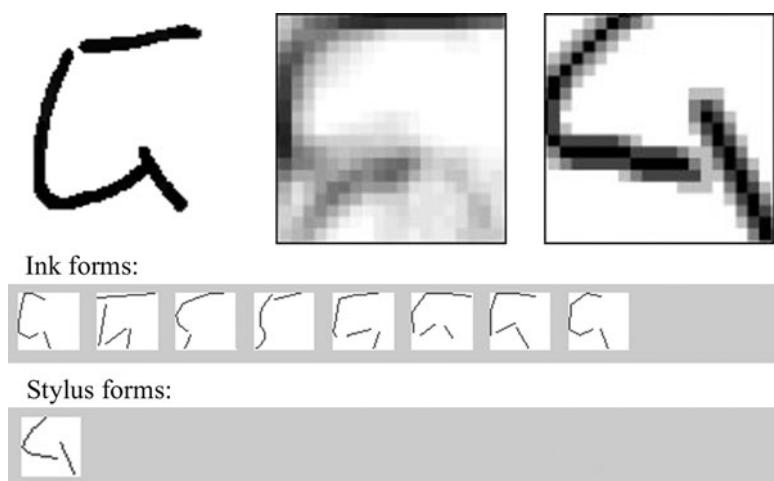
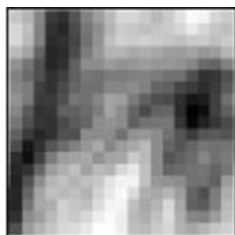


Figure C.7. The standard representation of the character G given in Bowman and Thomas (1983). The character model of the character G generated from all the ink characters in the training corpus, and the character model of the character G generated from all the stylus characters in the training corpus. Every instance of the character G in the ink tablet annotated corpus. The one instance of the character G in the stylus tablet annotated corpus.



Ink forms:

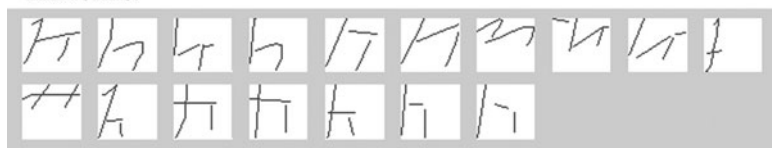
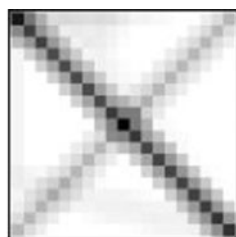
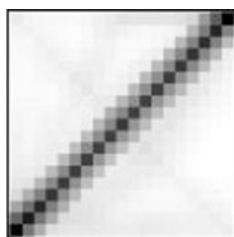
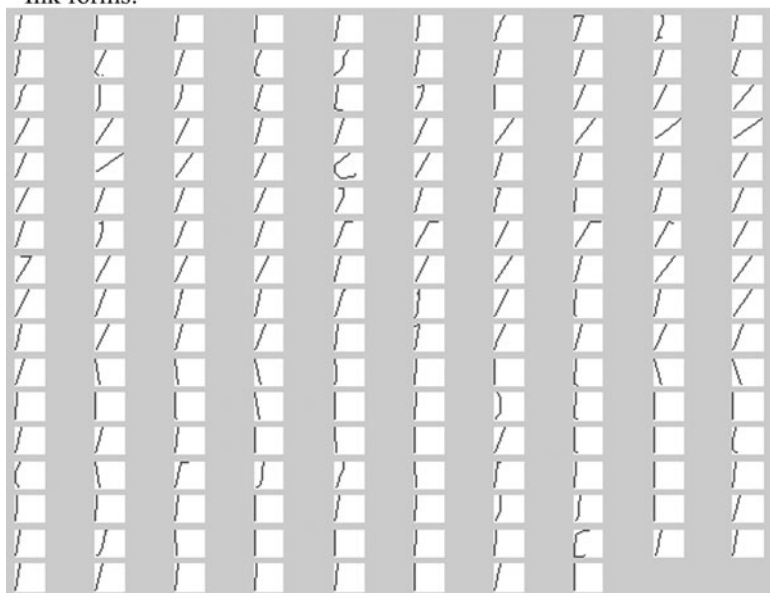


Figure C.8. The standard representation of the character H given in Bowman and Thomas (1983). The character model of the character H generated from all the ink characters in the training corpus. There were no instances of the character H in the training corpus of the annotated stylus tablets. Every instance of the character H in the ink tablet annotated corpus.



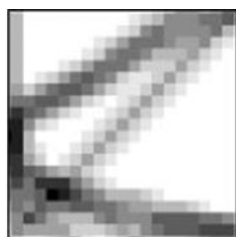
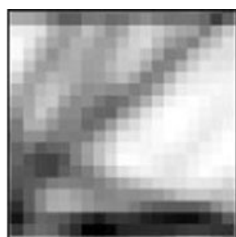
Ink forms:



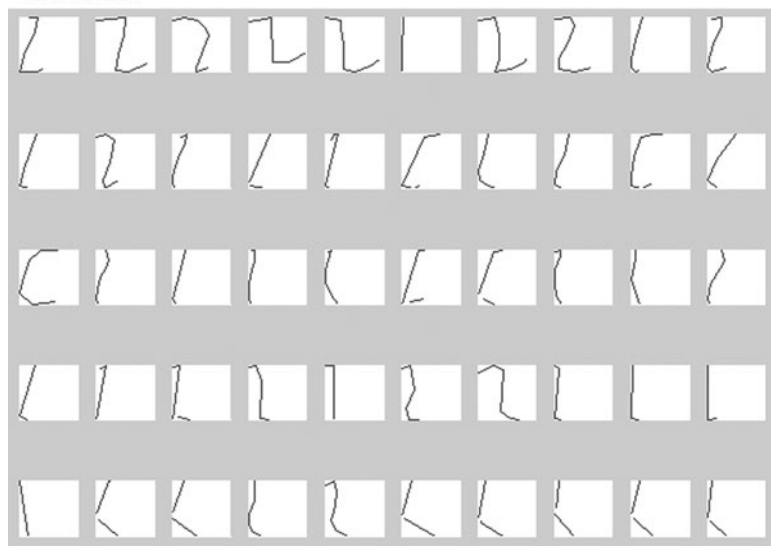
Stylus forms:



Figure C.9. The standard representation of the character I given in Bowman and Thomas (1983). The character model of the character I generated from all the ink characters in the training corpus, and the character model of the character I generated from all the stylus characters in the training corpus. Every instance of the character I in the ink tablet annotated corpus. Every instance of the character I in the stylus tablet annotated corpus.



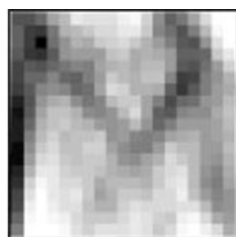
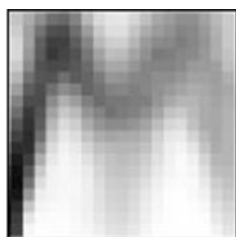
Ink forms:



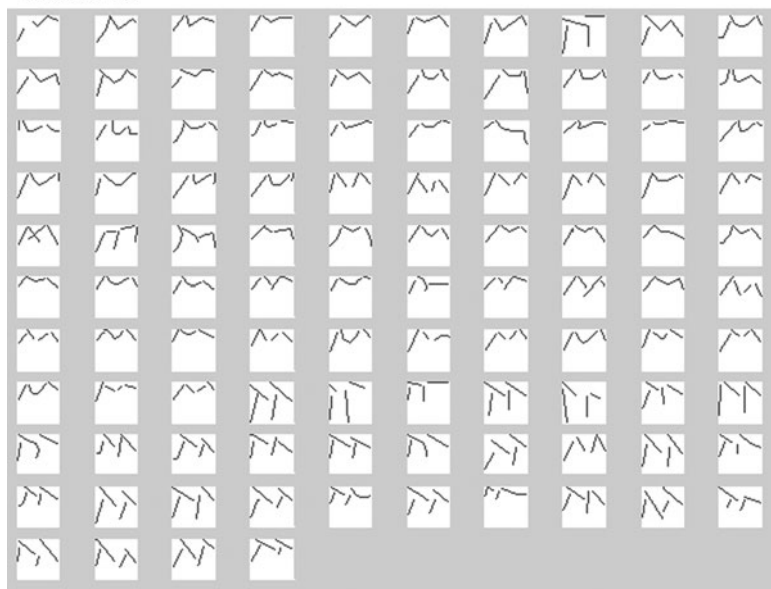
Stylus forms:



Figure C.10. The standard representation of the character L given in Bowman and Thomas (1983). The character model of the character L generated from all the ink characters in the training corpus, and the character model of the character L generated from all the stylus characters in the training corpus. Every instance of the character L in the ink tablet annotated corpus. Every instance of the character L in the stylus tablet annotated corpus.



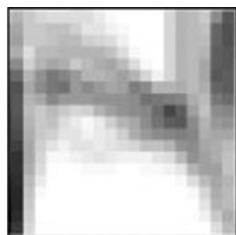
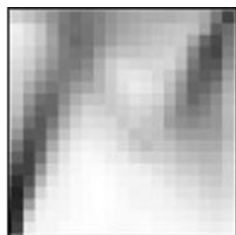
Ink forms:



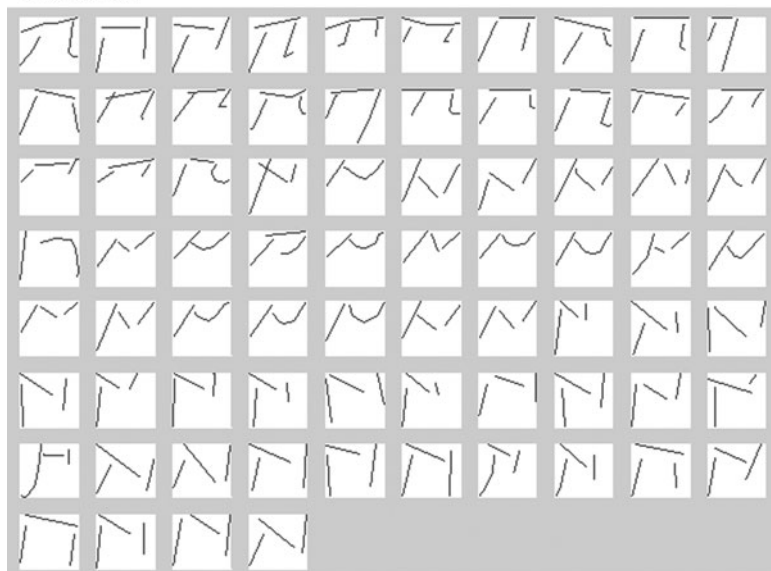
Stylus forms:



Figure C.11. The standard representation of the character M given in Bowman and Thomas (1983). The character model of the character M generated from all the ink characters in the training corpus, and the character model of the character M generated from all the stylus characters in the training corpus. Every instance of the character M in the ink tablet annotated corpus. Every instance of the character M in the stylus tablet annotated corpus.



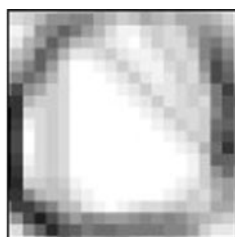
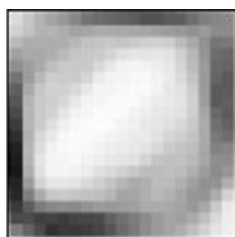
Ink forms:



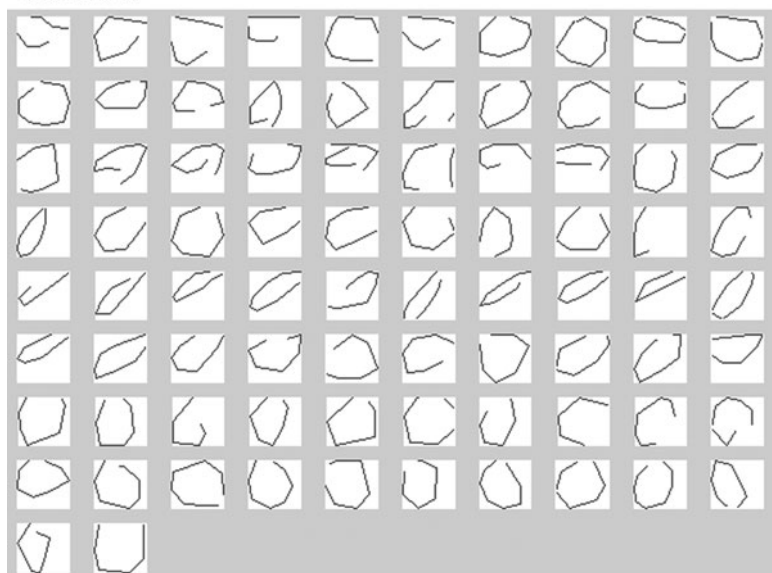
Stylus forms:



Figure C.12. The standard representation of the character N given in Bowman and Thomas (1983). The character model of the character N generated from all the ink characters in the training corpus, and the character model of the character N generated from all the stylus characters in the training corpus. Every instance of the character N in the ink tablet annotated corpus. Every instance of the character N in the stylus tablet annotated corpus.



Ink forms:



Stylus forms:



Figure C.13. The standard representation of the character O given in Bowman and Thomas (1983). The character model of the character O generated from all the ink characters in the training corpus, and the character model of the character O generated from all the stylus characters in the training corpus. Every instance of the character O in the ink tablet annotated corpus. Every instance of the character O in the stylus tablet annotated corpus.

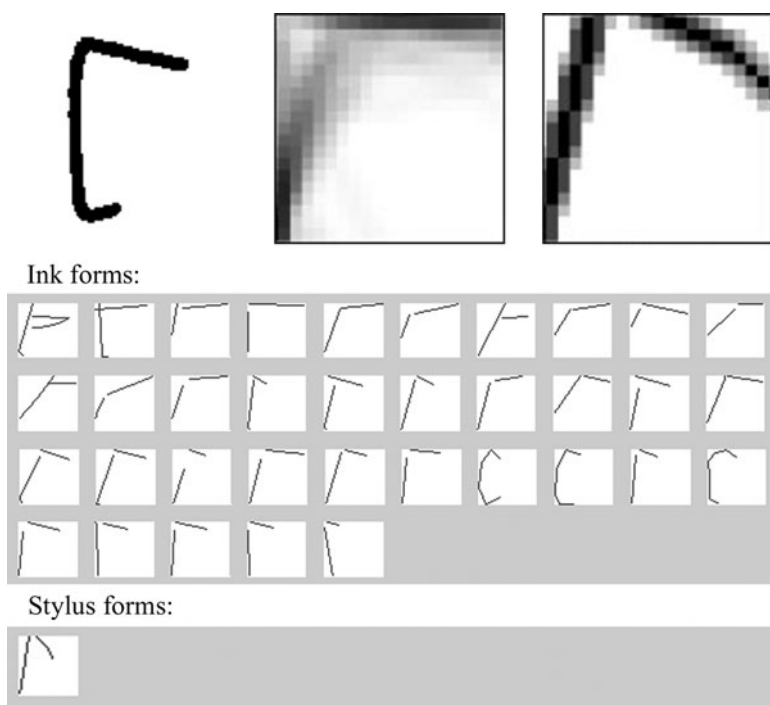


Figure C.14. The standard representation of the character P given in Bowman and Thomas (1983). The character model of the character P generated from all the ink characters in the training corpus, and the character model of the character P generated from all the stylus characters in the training corpus. Every instance of the character P in the ink tablet annotated corpus. Every instance of the character P in the stylus tablet annotated corpus.

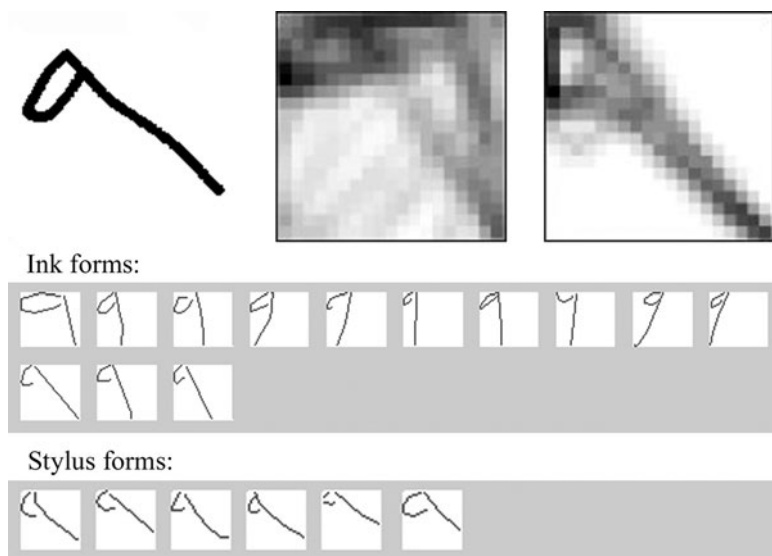


Figure C.15. The standard representation of the character Q given in Bowman and Thomas (1983). The character model of the character Q generated from all the ink characters in the training corpus, and the character model of the character Q generated from all the stylus characters in the training corpus. Every instance of the character Q in the ink tablet annotated corpus. Every instance of the character Q in the stylus tablet annotated corpus.

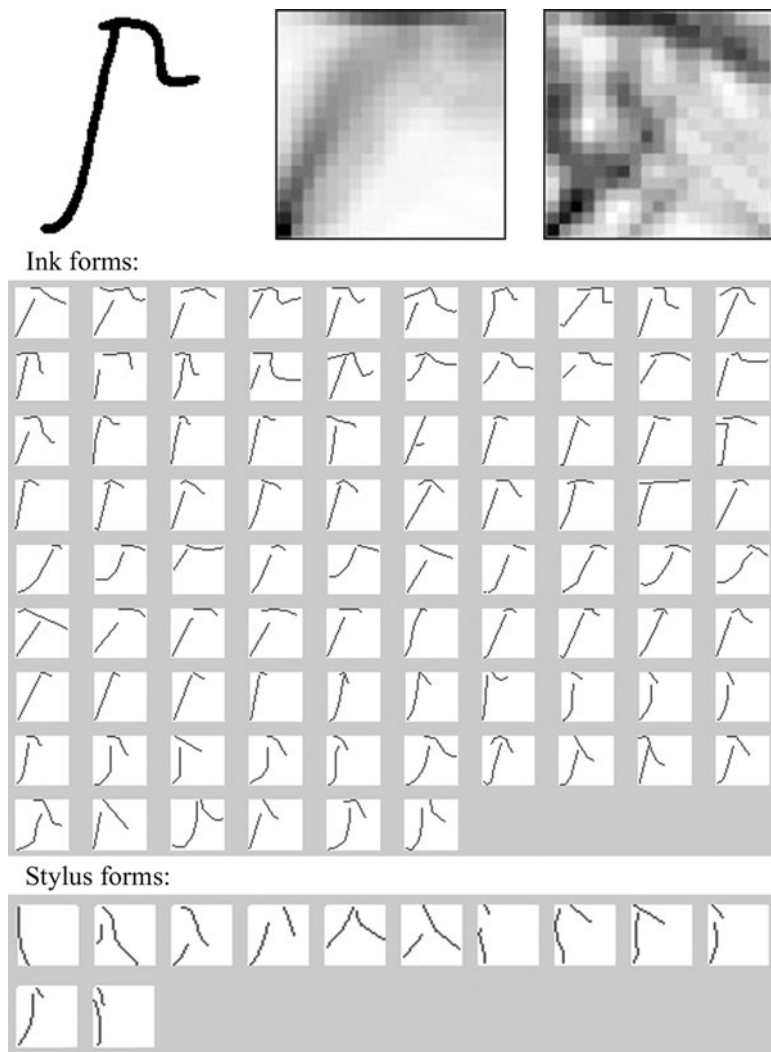
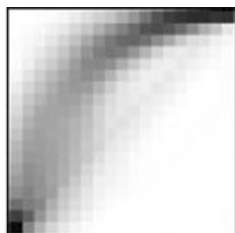
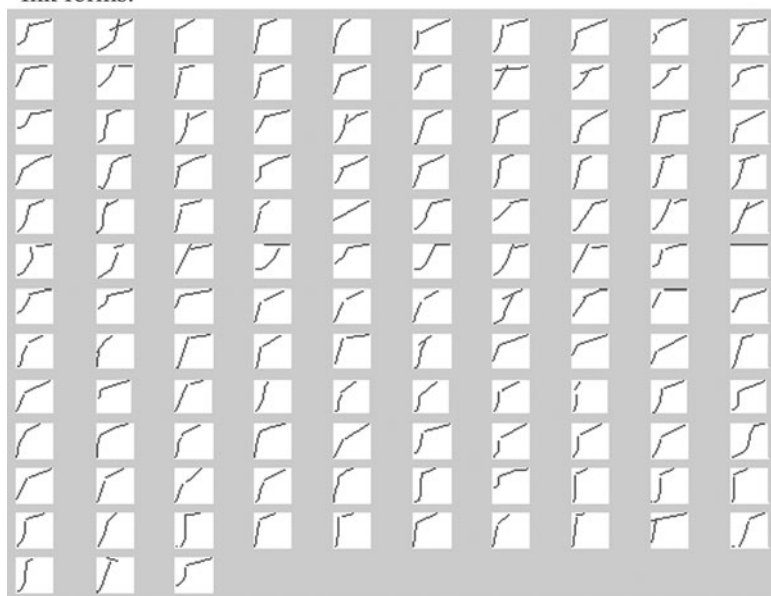


Figure C.16. The standard representation of the character R given in Bowman and Thomas (1983). The character model of the character R generated from all the ink characters in the training corpus, and the character model of the character R generated from all the stylus characters in the training corpus. Every instance of the character R in the ink tablet annotated corpus. Every instance of the character R in the stylus tablet annotated corpus.



Ink forms:



Stylus forms:



Figure C.17. The standard representation of the character S given in Bowman and Thomas (1983). The character model of the character S generated from all the ink characters in the training corpus, and the character model of the character S generated from all the stylus characters in the training corpus. Every instance of the character S in the ink tablet annotated corpus. Every instance of the character S in the stylus tablet annotated corpus.

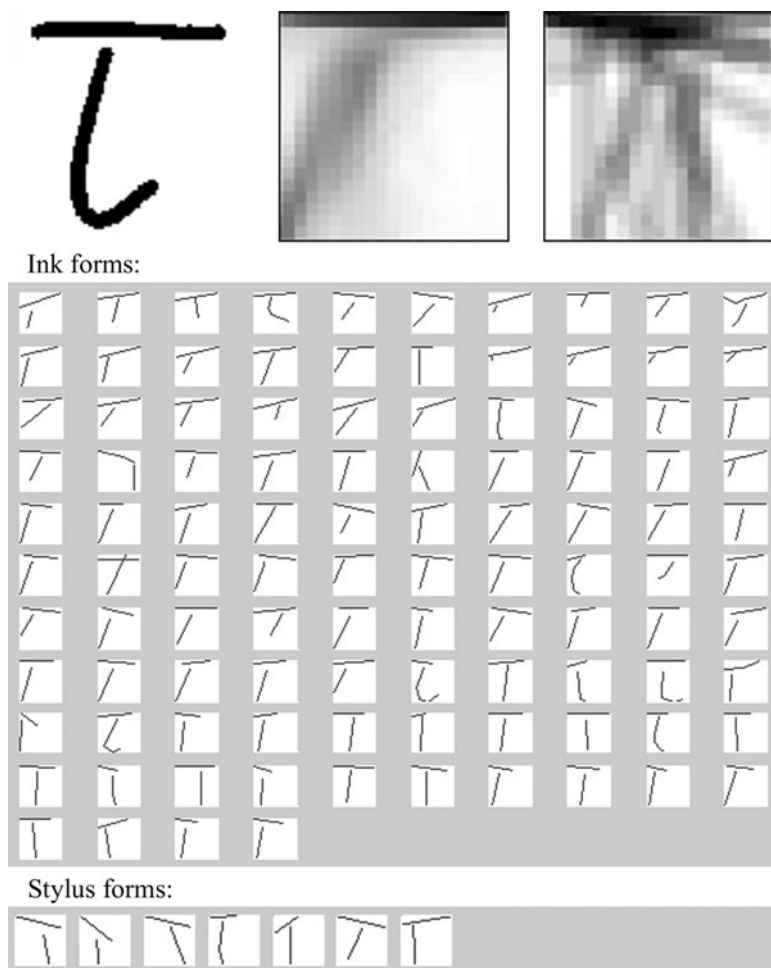
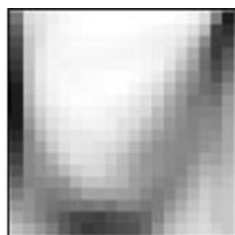
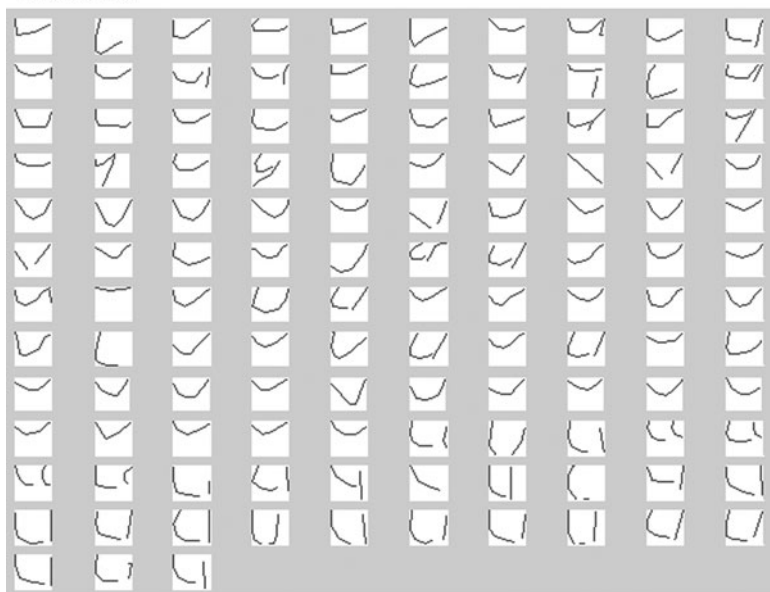


Figure C.18. The standard representation of the character T given in Bowman and Thomas (1983). The character model of the character T generated from all the ink characters in the training corpus, and the character model of the character T generated from all the stylus characters in the training corpus. Every instance of the character T in the ink tablet annotated corpus. Every instance of the character T in the stylus tablet annotated corpus.



Ink forms:



Stylus forms



Figure C.19. The standard representation of the character U given in Bowman and Thomas (1983). The character model of the character U generated from all the ink characters in the training corpus. There were no instances of the character U in the training corpus of the annotated stylus tablets, meaning no model was generated, but there were some examples in the test set, and these are shown below. Every instance of the character U in the ink tablet annotated corpus. Every instance of the character U in the stylus tablet annotated corpus (these were not present in the training set, however).

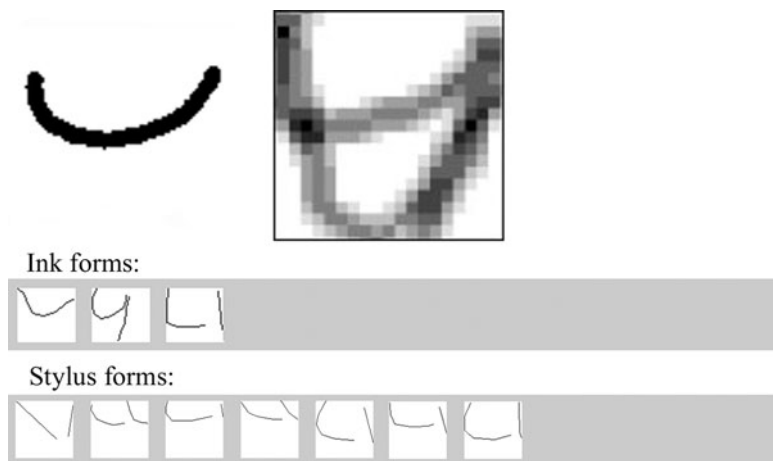


Figure C.20. The standard representation of the character V given in Bowman and Thomas (1983). The character model of the character V generated from all the ink characters in the training corpus. There were no instances of the character V in the training corpus of the annotated stylus tablets, meaning no model was generated, but there were some examples in the test set, and these are shown below. Every instance of the character V in the ink tablet annotated corpus. Every instance of the character V in the ink tablet annotated corpus (these were not present in the training set, however).

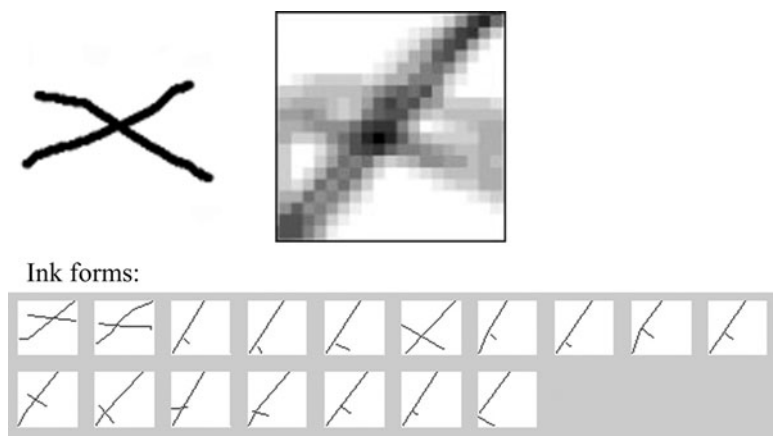


Figure C.21. The standard representation of the character X, provided by one of the experts in the Knowledge Elicitation exercises. The character model of the character X generated from all the ink characters in the training corpus. There were no instances of the character X in the training corpus, or the test corpus, of the annotated stylus tablets. Every instance of the character X in the ink tablet annotated corpus.

References

- Aalto, P. (1945). 'Notes on Methods of Decipherment of Unknown Writings and Languages', *Studia Orientalia*, *Edidat Societas Orientalis Fennica*, 11(4).
- Adams, J. N. (1995). 'The Language of the Vindolanda Writing Tablets: An Interim Report', *Journal of Roman Studies*, 85: 86–134.
- M. Janse, and S. Swain, eds. (2002). *Bilingualism in Ancient Society: Language, Contact and the Written Word* (Oxford: Oxford University Press).
- Aiolfi, F. (1999). 'Sviluppo di un sistema basato sulla Distanza Tangente per l'Analisi Morfologica di Caratteri Paleografici', Tesi di Laurea (masters thesis), Department of Computer Science, University of Pisa, Feb.
- M. Simi, D. Sona, A. Sperduti, A. Starita, and G. Zaccagnini (1999). 'SPI: A System for Paleographic Inspections', *Workshop on Intelligenza Artificiale per i Beni Ambientali (AI*IA)*, Notiziario AI*IA, 4 (Dec). <http://sra.itc.it/people/sona/PAPERS/aiolfi99spi.pdf> (accessed 1 April 2005).
- Aiken, L. R. (1971). *Psychological Testing and Assessment* (London: Allyn & Bacon).
- Aitchison, J. (1998). *The Articulate Mammal: An Introduction to Psycholinguistics* (London: Routledge).
- Anderson, B. (1974). *The Quantifier as Qualifier: Some Notes on Qualitative Elements in Quantitative Content Analysis* (Gothenburg: University of Gothenburg).
- André, J. (1981). *L'Alimentation et la cuisine à Rome* (Paris: Belles Lettres).
- (1985). *Les Noms de plantes dans la Rome antique* (Paris: Belles Lettres).
- (1991). *Le Vocabulaire latin de l'anatomie* (Paris: Belles Lettres).
- Asher, R. E., ed. (1994). *The Encyclopedia of Language and Linguistics* (Oxford: Pergamon).
- Bagnall, R. S. (1997). 'Imaging of Papyri: A Strategic View', *Literary and Linguistic Computing*, 12(3): 153–4.
- Bately, J., M. P. Brown, and J. Roberts, eds. (1993). *A Palaeographer's View: The Selected Writings of Julian Brown* (London: Harvey Miller Publishers).
- Baum L. E. (1972). 'An Inequality and Associated Maximization Technique in Statistical Estimation for Probabilistic Functions of a Markov Process', *Inequalities*, 3: 1–8.

- Bearman, G. H., and S. Spiro (1996). 'Archaeological Applications of Advanced Imaging Techniques', *Biblical Archaeologist*, 59(1): 56–66.
- S. J. Pfann, and S. Spiro (1998). 'Imaging the Scrolls: Photographic and Direct Digital Acquisition', in P. W. Flint and J. C. VanderKam (eds.), *The Dead Sea Scrolls after Fifty Years: A Comprehensive Assessment* (Leiden: Brill Academic Publishers), i. 472–95.
- Benton, J. F., A. R. Gillespie, and J. M. Soha (1979). 'Digital Image-Processing Applied to the Photography of Manuscripts, with Examples Drawn from the Pincus ms. of Arnald of Villanova', *Scriptorium*, 33: 40–55.
- Biber, D. (1983). 'Representativeness in Corpus Design', *Literary and Linguistic Computing*, 8(4): 243–57.
- (1990). 'Methodological Issues Regarding Corpus-Based Analysis of Linguistic Variation', *Literary and Linguistic Computing*, 5(4): 257–69.
- Bidwell, P. T. (1985). *The Roman Fort of Vindolanda at Chesterholm, Northumberland* (London: Historic Buildings and Monuments Commission for England).
- (1997). *Roman Forts in Britain* (Bath: English Heritage, The Bath Press).
- Bingen, J., A. Bülow-Jacobsen, W. E. H. Cockle, H. Cuvigny, F. Kayser, and W. Van Rengen (1997). *Mons Claudianus, Ostraca Graeca et Latina II* (O.Claud. 191–414 (DFIFAO XXXII); Cairo: Institut français d'archéologie orientale du Caire).
- Birley, E., R. Birley, and A. R. Birley (1993). *The Early Wooden Forts: Reports on the Auxiliaries, the Writing Tablets, Inscriptions, Brands and Graffiti* (Hexham: Roman Army Museum Publications for the Vindolanda Trust).
- Birley, R. E. (1977). *Vindolanda: A Roman Frontier Post on Hadrian's Wall* (London: Thames & Hudson).
- (1997). 'The Vindolanda Bonfire', *Current Archaeology*, 13(9): 348–57.
- (1999). 'Vindolanda IV: Fasc. IV', *Writing Materials* (Greenhead, Carvoran, Northumberland: Roman Army Museum Publications).
- Bischoff, B. (1990). *Latin Palaeography: Antiquity and the Middle Ages*, tr. D. O. Croinin and D. Ganz (Cambridge: Cambridge University Press).
- Booras, S. W., and D. M. Chabries (2001). 'The Herculaneum Scrolls', *Proceedings of Image Processing, Image Quality, Image Capture Systems Conference (PICS-01)* (Montreal and Quebec: Society for Imaging Science and Technology), 215–18.
- Boose, J. H. (1990). 'Uses of Repertory Grid-Centered Knowledge Acquisition Tools for Knowledge-Based Systems', in J. Boose and B. Gaines, *Foundations of Knowledge Acquisition* (London: Academic Press), 61–83.
- and B. Gaines, eds. (1990). *The Foundation of Knowledge Acquisition* (London: Academic Press).

- Boswinkel, E., and P. W. Pestman, eds. (1978). *Textes grecs, demotiques et bilingues* (Papyrologica Lugduno-Batava, 19; Leiden: Brill Academic Publishers).
- Bowman, A. K. (1994) *Life and Letters on the Roman Frontier* (London: British Museum Press).
- (1999). 'The Vindolanda Writing-Tablets 1991–4', *Atti del XI Congresso Internazionale di Epigrafia Greca e Latina, Roma 18–24 Settembre 1997* (Rome: Edizion: Quasar), i. 545–51.
- (2003). *Life and Letters on the Roman Frontier: Vindolanda and its People*, 2nd edn. (London: British Museum Press).
- and M. Brady, eds. (2005). *Images and Artefacts of the Ancient World* (Oxford: Oxford University Press for British Academy).
- and J. D. Thomas (1974). *The Vindolanda Writing Tablets* (Newcastle upon Tyne: Graham).
- and —— (1983). *Vindolanda: The Latin Writing Tablets (Tab. Vindol. I)* (London: Society for Promotion of Roman Studies).
- and —— (1994). *The Vindolanda Writing-Tablets (Tab. Vindol. II)* (London: British Museum Press).
- and —— (2003). *The Vindolanda Writing-Tablets (Tab. Vindol. III)* (London: British Museum Press).
- J. M. Brady, and R. S. O. Tomlin (2005). 'Wooden Stylus Tablets from Roman Britain', in A. K. Bowman and M. Brady (eds.), *Images and Artefacts of the Ancient World* (Oxford: Oxford University Press), 7–14.
- J. M. Brady, and R. S. O. Tomlin (1997). 'Imaging Incised Documents', *Literary and Linguistic Computing*, 12(3): 169–76.
- Boyle, L. E. (2001). *Integral Palaeography* (Turnhout: Brepolis).
- Brady, M., X. Pan, M. Terras, and V. Schenk (2005). 'Shadow Stereo, Image Filtering and Constraint Propagation', in A. K. Bowman and M. Brady (eds.), *Images and Artefacts of the Ancient World* (Oxford: Oxford University Press), 15–30.
- Breeze, D. J., and B. Dobson (1976). *Hadrian's Wall* (London: Richard Clay (The Chaucer Press)).
- Brooks, R. A. (1986). 'A Robust Layered Control System for a Mobile Robot', *Institute of Electrical and Electronics Engineers Journal of Robotics and Automation*, RA-2: 14–23.
- Brown, M. S., and W. B. Seales (2001). *3D Imaging and Processing of Damaged Texts*, Joint International Conference of the Association for Computers and the Humanities and the Association for Literary and Linguistic Computing, New York University, 13–16 June. http://www.nyu.edu/its/humanities/ach_allc2001/papers/brown/ (accessed 27 Jan 2005).

- Burrus, C. S., R. A. Gopinath, and H. Guo (1997). *Introduction to Wavelets and Wavelet Transforms: A Primer* (Upper Saddle River, NJ: Prentice Hall).
- Carr, M., J. Durand, and S. Thomas (1994). *Translation Theory and Practice: A Discourse* (Salford: University of Salford).
- Carr, T., and B. A. Levy, eds. (1990). *Reading and its Development: Component Skills Approaches* (London: Academic Press).
- Casamassima, E., and E. Staraz (1977). 'Varianti e cambio grafico nella scrittura dei papiri latini: note palaeografiche', *Scrittura e civiltà*, 1: 9–110.
- Cattel, J. M. (1886). 'The Time Taken up by Cerebral Operations', *Mind*, 2: 220–42.
- Cencetti, G. (1950). *Note Paleografiche sulla Scrittura dei Papiri Latini dal I al III Secolo D.C.* (Memorie dell'Accademia delle scienze dell'Istituto di Bologna. Classe di scienze morali, V/1).
- Chabries, D. M., and S. W. Booras (2001). 'The Petra Scrolls', *Proceedings of Image Processing, Image Quality, Image Capture Systems Conference (PICS-01)* (Montreal and Quebec: Society for Imaging Science and Technology), 198–295.
- and — and G. H. Bearman (2003). 'Imaging the Past: Recent Applications of Multispectral Imaging Technology to Deciphering Manuscripts', *Antiquity*, 77: 359–72.
- Charniak, E. (1993). *Statistical Language Learning* (Cambridge, Mass.: Massachusetts Institute of Technology Press).
- Chomsky, N. (1957). *Syntactic Structures* (The Hague: Mouton).
- (1972). *Language and Mind* (New York: Harcourt Brace Jovanovich).
- Ciula, A. (2004). 'A Research Project: The Application of SPI Software to the Corpus of Manuscripts Held in Siena', MA dissertation, Master of Arts in Applied Computing in the Humanities, King's College London.
- Clanchy, M. T. (1993). *From Memory to Written Record, England 1066–1307* (Oxford: Blackwell).
- Coleman, R. G. G. (1993). 'Vulgar Latin and Proto Romance: Minding the Gap', *Prudentia*, 25(2): 1–14.
- Connell, J. H., and M. Brady (1987). 'Generating and Generalizing Models of Visual Objects', *Artificial Intelligence*, 31(2): 159–83.
- Connell, J. L., and L. Shafer (1989). *Structured Rapid Prototyping: An Evolutionary Approach to Software Development* (Upper Saddle River, NJ: Yourdon Press Computing Series).
- Conway, R. S. (1923). *An Introduction to Latin, Greek and English Etymology* (London: John Murray).
- Cordingley, E. (1989). 'Knowledge Elicitation Techniques for Knowledge Based Systems', in D. Diaper (ed.), *Knowledge Elicitation: Principles, Techniques and Applications* (Chichester: Ellis Horwood), 90–103.

- Coulmas, F. (1989). *The Writing Systems of the World* (Oxford: Blackwell).
- Côté, M., E. Lecolinet, M. Cheriet, and C. Y. Suen. (1998). 'Automatic Reading of Cursive Scripts Using a Reading Model and Perceptual Concepts: The PERCEPTO System', *International Journal of Document Analysis and Recognition*, 1(1): 3–17.
- Crain, S., and M. Steedman (1985). 'On Not Being Led up the Garden Path: The Use of Context by the Psychological Syntax Processor', in D. Dowty, L. Karttunen, and A. M. Zwicky (eds.), *Natural Language Parsing: Psycholinguistic, Computational and Theoretical Perspectives* (Studies in Natural Language Processing; Cambridge: Cambridge University Press), 320–58.
- Dahl, R. (1997). *The Great Automatic Grammatizator and Other Stories* (London: Penguin Books).
- Danks, J. H., and J. Griffin (1997). 'Reading and Translation: A Psycholinguistic Perspective', in J. H. Danks, G. M. Shreve, S. B. Fountain, and M. K. McBeath (eds.), *Cognitive Processes in Translation and Interpreting* (Thousand Oaks, Calif.: Sage Publications), 161–75.
- G. M. Shreve, S. B. Fountain, and M. K. McBeath, eds. (1997). *Cognitive Processes in Translation and Interpreting* (Thousand Oaks, Calif.: Sage Publications).
- De Carteret, C., and R. Vidgen (1995). *Data Modelling for Information Systems* (London: Pitman).
- De Cervantes, M. (1615). *Don Quixote*, tr. E. Grossman, 2004 (London: Secker & Walburg).
- De Groot, A. M. B. (1997). 'The Cognitive Study of Translation and Interpretation', in J. H. Danks, G. M. Shreve, S. B. Fountain, and M. K. McBeath (eds.), *Cognitive Processes in Translation and Interpreting* (Thousand Oaks, Calif.: Sage Publications).
- and C. Barry (1994). *The Multilingual Community: Bilingualism* (Hillsdale, NJ: Lawrence Erlbaum Associates).
- Dermott, J. (1988). 'Preliminary Steps toward a Taxonomy of Problem Solving Methods', in S. Marcus (ed.), *Automating Knowledge Acquisition for Expert Systems* (Lancaster: Kluwer Academic Publishers), 225–56.
- Diaper, D. (1989). *Knowledge Elicitation: Principles, Techniques and Applications* (Chichester: Ellis Horwood).
- Dodel, J. P., and R. Shinghal (1995). 'Symbolic/Neuronal Recognition of Cursive Amounts on Bank Cheques', *Proceedings of the Third International Conference on Document Analysis and Recognition*, Montreal, i. 15–18.
- Dunker, K. (1926). 'A Qualitative (Experimental and Theoretical) Study of Productive Thinking (Solving of Comprehensible Problems)', *Pedagogical Seminar*, 33: 642–708.

- Ellegard, A. (1963). *English, Latin, and Morphemic Analysis* (Gothenburg: University of Gothenburg).
- Ellis, R., and G. Humphreys (1999). *Connectionist Psychology: A Text with Readings* (Hove: Psychology Press).
- Ericsson, K. A., and H. A. Simon (1993). *Protocol Analysis: Verbal Reports as Data* (Cambridge, MA: MIT Press).
- Erman, L. D., F. Hayes-Roth., V. R. Lesser, and D. R. Reddy (1980). 'The HEARSAY-II Speech Understanding System: Integrating Knowledge to Resolve Uncertainty', *Computing Surveys*, 12(2): 213–53.
- Eysenck, M. W., and M. T. Keane (1997). *Cognitive Psychology: A Student's Handbook* (Hove: Psychology Press).
- Fahlman, S. E. (1973), 'A Planning System for Robot Construction Tasks', *Technical Report 283*, Massachusetts Institute of Technology Artificial Intelligence Lab, <http://hdl.handle.net/1721.1/6918> (accessed 27 Jan 2005).
- Feigenbaum, E. A. (1977). *The Art of Artificial Intelligence: Themes and Case Studies in Knowledge Engineering* (Stanford, Calif.: International Joint Conference of Artificial Intelligence, 5; 1977); 1014–29.
- Fink, R. O. (1971). *Roman Military Records on Papyrus* (Philological Monographs of the American Philological Association, 26; Cleveland, Oh.: Case Western Reserve University Press).
- Finney, J. T. (1999). 'The Ancient Witnesses of the Epistle to the Hebrews: A Computer-Assisted Analysis of the Papyrus and Uncial Manuscripts of Pros Ebraious', D.Phil. thesis, Murdoch University, Perth, Australia, <http://alpha.reltech.org/TFinney/thesis/index.html> (accessed 31 Jan 2005).
- Forster, M. R. (1999). 'The New Science of Simplicity', in H. A. Keuzenkamp, M. McAleer, and A. Zellner (eds.), *Simplicity, Inference and Econometric Modelling* (Cambridge: Cambridge University Press), 83–117.
- Francis, W. N., and H. Kucera (1979). *Brown Corpus Manual* (Providence, Brown University), <http://helmer.aksis.uib.no/icame/brown/bcm.html> (accessed 25 Feb 2005).
- Franzosi, R. (1990). 'Strategies for the Prevention, Detection and Correction of Measurement Error in Data Collected from Textual Sources', *Sociological Methods and Research*, 18: 442–71.
- (1997). 'Labor Unrest in the Italian Service Sector: An Application of Semantic Grammars', in C. W. Roberts (ed.), *Text Analysis for the Social Sciences* (Mahwah, NJ: Lawrence Erlbaum Associates), 131–45.
- Fu, K. S., and P. H. Swain (1969). 'On Syntactic Pattern Recognition', in J. T. Tou (ed.), *Software Engineering*, 2: 155–82.
- Furneaux, S., and M. F. Land (1997). 'The Role of Eye Movements during Music Reading', *Proceedings of the Third European Society for the Cognitive*

- Sciences of Music (ESCOM) Conference* (Uppsala: Uppsala University), 210–14.
- Gaiman, N. (2003). 'In San Diego', *Neilgaiman.com Journal Entry*, 17 July 2003, http://www.neilgaiman.com/journal/2003_07_13_archive.asp (accessed 3 June 2004).
- Gaines, B. R., and M. L. G. Shaw (1997). *WebGrid: Knowledge Modeling and Inference through the World Wide Web* (Knowledge Science Institute, University of Calgary), <http://repguid.com/reports/KBS/KMD/> (accessed 5 June 2002).
- Gallo, I. (1986). *Greek and Latin Papyrology* (London: Institute of Classical Studies).
- Gao, Q., and L. Ming (2000). 'Applying MDL to Learning Best Model Granularity', *Artificial Intelligence*, 121: 1–29.
- Gaur, A. (1987). *A History of Writing* (London: British Library Publications).
- Geman, S., and D. Geman (1984). 'Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images', *Institute of Electrical and Electronics Engineers Transactions on Pattern Analysis and Machine Intelligence*, 6: 721–41.
- Gibson, E. J., and H. L. Levin (1976). *The Psychology of Reading* (Cambridge, Mass.: MIT Press).
- Gonzalez, R. C., and R. E. Woods (1993). *Digital Image Processing* (Reading, Mass.: Addison-Wesley Publishing).
- Goodman, K. S. (1967). 'Reading: A Psycholinguistic Guessing Game', *Journal of the Reading Specialist*, 6: 126–35.
- Gough, P. B. (1972). 'One Second of Reading', in J. F. Kavanagh and I. G. Mattingly (eds.), *Language by Ear and by Eye* (Cambridge, Mass.: MIT Press).
- and M. Cosky (1977). 'One Second of Reading Again', in N. J. Castellan, D. B. Pisoni, and G. R. Potts, *Cognitive Theory (2)* (Hillsdale, N.J.: Lawrence Erlbaum Associates).
- Gregory, R. L. (1994). *Eye and Brain: The Psychology of Seeing* (Oxford: Oxford University Press).
- Grenfell, B. P. (1918). 'New Papyri from Oxyrhynchus', *Journal of Egyptian Archaeology*, 5(1): 18–23.
- (1921). 'The Present Position of Papyrology', *Bulletin of the John Rylands Library*, 6 (1–2): 1–21.
- Hall, F. W. (1913). *A Companion to Classical Texts* (Oxford: Clarendon Press).
- Hammersley, J. M., and D. C. Handscomb (1964). *Monte Carlo Methods* (London: Chapman & Hall).
- Hart-Davis, R., ed. (1978), *The Lyttelton Hart-Davis Letters*, i. (London: Murray).

- Healey, J. F. (1990). *The Early Alphabet* (London: British Library Publications).
- Henderson, J. M., and A. Hollingworth (1998). 'Eye Movements during Scene Viewing: An Overview', in G. Underwood (ed.), *Eye Guidance in Reading and Scene Perception* (Oxford: Elsevier), 269–93.
- Herman, J. (2000). *Vulgar Latin*, tr. T. B. R. Wright (Philadelphia: Pennsylvania State University Press).
- Hilton, O. (1959). 'Characteristics of the Ball Point Pen and its Influence on Handwriting Identification', *Journal of Criminal Law, Criminology, and Police Science*, 47: 606–13.
- (1984). 'Effects of Writing Instruments on Handwriting Details', *Journal of Forensic Sciences*, 29: 80–6.
- Hippisley, D. (2005) 'Encoding the Vindolanda Tablets: An Investigation in Contextual Encoding using XML and the EpiDoc Standards', MA dissertation, Masters in Electronic Communication and Publishing, University College London.
- Hirst, G. (1987). *Semantic Interpretation and the Resolution of Ambiguity* (Cambridge: Cambridge University Press).
- Hockey, S. (2000). *Electronic Texts in the Humanities* (Oxford: Oxford University Press).
- Hoey, M. (2001). *Textual Interaction: An Introduction to Written Discourse Analysis* (London: Routledge).
- Hogaboam, T. W., and C. A. Perfetti (1975). 'Lexical Ambiguity and Sentence Comprehension', *Journal of Verbal Learning and Verbal Behaviour*, 16(3): 265–74.
- Holsti, O. R. (1969). *Content Analysis for the Social Sciences and Humanities* (Reading, Mass.: Addison-Wesley).
- Hornby, P. A. (1977). *Bilingualism, Psychological, Social and Educational Implications* (New York: Academic Press).
- Holy Bible, King James VI's Version* (1953). (Glasgow, Collins' Clear-type Press).
- Huhns, M. N., and M. P. Singh, eds. (1998). *Readings in Agents* (San Francisco: Morgan Kaufmann Publishers).
- Hunt, A. S. (1914). 'Papyri and Papyrology', *Journal of Egyptian Archaeology*, 1(2): 81–91.
- (1922). 'Twenty-Five Years of Papyrology', *Journal of Egyptian Archaeology*, 8 (3–4): 121–8.
- (1928). 'Review Article: Greek Papyri in the Library of Cornell University', *Journal of Egyptian Archaeology*, 14 (1–2): 24–5.
- (1930). 'Papyrology', *Oxford Magazine* (30 Oct.): 84–6.

- Illingworth, V., ed. (1996). *The Oxford Dictionary of Computing*, 4 edn. (Oxford: Oxford University Press).
- Impedovo, S. (1993). *Fundamentals in Handwriting Recognition* (NATO Advanced Study Workshop on Fundamentals in Handwriting Recognition; Château de Bonas, France: Springer-Verlag).
- Ireland, S. (1986). *Roman Britain: A Sourcebook* (London: Routledge).
- Jameson, M. (2004). 'Promises and Challenges of Digital Libraries and Document Image Analysis: A Humanist's Perspective', *Proceedings of the 2004 International Workshop on Document Image Analysis for Libraries*, Institute of Electrical and Electronics Engineers, <http://csdl.computer.org/comp/proceedings/dial/2004/2088/00/2088toc.htm> (accessed 4 May 2004).
- Karssemeijer, N. (1992). 'Stochastic Model for Automated Detection of Calcifications in Digital Mammograms', *Image and Vision Computing*, 10 (6): 369–75.
- Kelly, G. A. (1955). *The Psychology of Personal Constructs*, (New York: W. W. Norton & Co.).
- Kenny, A. (1982). *The Computation of Style: An Introduction to Statistics for Students of Literature and Humanities* (Oxford: Pergamon Press).
- Kenyon, F. G. (1918). *Greek Papyri and their Contribution to Classical Literature: A Paper Communicated to the Leeds Branch [of the Classical Association] at its Fourth Annual Meeting 26 January 1918* (Cambridge: Cambridge University Press).
- Kiernan, K. S. (1991). 'Digital Image Processing and the Beowulf Manuscript', *Literary and Linguistic Computing*, 6: 20–7.
- Kiraly, D. C. (1997). 'Think Aloud Protocols and the Construction of a Professional Translator Self-Concept', in J. H. Danks, G. M. Shreve, S. B. Fountain, and M. K. McBeath (eds.), *Cognitive Processes in Translation and Interpreting* (Thousand Oaks, Calif.: Sage Publications), 137–60.
- Kleve, K., E. S. Ore, and R. Jensen (1987). 'Literalogy: On the Use of Computer Graphics and Photography in Papyrology', *Symbolae Osloenses*, 62: 109–29.
- Knox, K., R. Johnston, and R. L. Easton (1997). 'Imaging the Dead Sea Scrolls', *Optics and Photonics News*, 8(8): 30–4.
- Krippendorff, K. (1980). *Content Analysis: An Introduction to its Methodology* (London: Sage Publications).
- Laddaga, R., and P. Robertson (1996). *Yolambda Reference Manual* (Haverhill, Mass.: Dynamic Object Language Labs, Inc.).
- Lalou, É., ed. (1992). *Les Tablettes à écrire de l'Antiquité à l'époque moderne* (Turnhout: Brepolis).
- Lanterman, A., M. I. Miller, and D. L. Snyder (1998). 'Minimum Description Length Understanding of Infrared Scenes', *Automatic Target Recognition VIII* (Orlando, Fla.: SPIE Proc. 3371), 375–86.

- Lau, T. W. E., and Y. C. Ho (1997). 'Universal Alignment Probabilities and Subset Selection for Ordinal Optimization', *Journal of Optimization Theory and Applications*, 93: 455–89.
- Lawler, J. M., and H. A. Dry, eds. (1998). *Using Computers in Linguistics: A Practical Guide* (London: Routledge).
- Leclerc, Y. G. (1989). 'Constructing Simple Stable Descriptions for Image Partitioning', *International Journal of Computer Vision*, 3(1): 73–102.
- Levison, M. (1965). 'The Siting of Fragments', *Computer Journal*, 7: 275–7.
- (1967). 'The Computer in Literary Studies', in A. D. Booth (ed.). *Machine Translation* (Amsterdam: North Holland Publishing Co.) 173–4.
- (1999). 'The Jigsaw Puzzle Problem Revisited', Joint International Conference of the Association for Computers and the Humanities and the Association for Literary and Linguistic Computing, Charlottesville, Va., ALLC.
- Liversedge, S. P., K. B. Paterson, and M. J. Pickering (1998). 'Eye Movements and Measures of Reading Time', in G. Underwood (ed.), *Eye Guidance in Reading and Scene Perception* (Oxford: Elsevier), 55–75.
- Lundberg, M. J. (2002). 'New Technologies: Reading Ancient Inscriptions in Virtual Light', West Semitic Research Project, University of Southern California, <http://www.usc.edu/dept/LAS/wsrp/information/article.html> (accessed 7 Sept 2002).
- McClelland, J. L., and D. E. Rumelhart (1981). 'An Interactive Activation Model of Context Effects in Letter Perception: Part 1. An Account of Basic Findings', *Psychological Review*, 88: 375–407.
- and —— (1986a). 'A Distributed Model of Human Learning and Memory', in J. L. McClelland and D. E. Rumelhart and the PDP Research Group, *Parallel Distributed Processing, ii. Psychological and Biological Models* (Cambridge, Mass.: MIT Press).
- and —— (1986b). 'The Programmable Blackboard Model of Reading', in D.E. Rumelhart, J. L. McClelland, and the PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, i. Foundations* (Cambridge, Mass.: MIT Press).
- and —— and the PDP Research Group (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, ii. Psychological and Biological Models* (Cambridge, Mass.: MIT Press).
- McDermott, D. V., and G. L. Sussman (1973). 'The Conniver Reference Manual', *Artificial Intelligence Memo 259* (Cambridge, Mass.: MIT Artificial Intelligence Lab).
- McGraw, K. L., and K. Harbison-Briggs (1989). *Knowledge Acquisition: Principles and Guidelines* (London: Prentice-Hall International Editions).

- McGraw, K. L., and C. R. Westphal, eds. (1990). *Readings in Knowledge Acquisition: Current Practices and Trends* (London: Ellis Horwood).
- Mahoney, A. (2000). 'Overview of Latin Syntax', Perseus Project, <http://www.perseus.tufts.edu/cgi-bin/ptext?doc=1999.04.0022> (accessed 6 June 2004).
- and J. Rydberg-Cox (2001). 'A Note on Scrabble in Latin', *Classical Outlook*, 78(2): 58–9.
- Mallon, J. (1952). *Paléographie Romaine* (Madrid: Consejo Superior de Investigaciones Científicas, Instituto Antonio de Nebrija, de Filología).
- Manguel, A. (1997). *A History of Reading* (London: Flamingo).
- Marcus, S., ed. (1988). *Automating Knowledge Acquisition for Expert Systems* (Lancaster: Kluwer Academic Press).
- Marichal, R. (1950). 'L'Écriture latine du Ier au VIIIe siècle: Les sources', *Scriptorium*, 4: 131–3.
- (1955). 'L'Écriture latine du Ier au VIIIe siècle: Les sources', *Scriptorium*, 9: 129–30.
- Masson, J. (1985). 'Felt Tip Pen Writing', *Journal of Forensic Sciences*, 30: 172–7.
- Mathyer, J. (1969). 'Influence of Writing Instruments on Handwriting and Signatures', *Journal of Criminal Law, Criminology and Police Science*, 60: 102–12.
- Millet, M. (1995). *Roman Britain* (Bath: English Heritage, The Bath Press).
- Molton, N., X. Pan, M. Brady, A. K. Bowman, C. Crowther, and R. Tomlin (2003). 'Visual Enhancement of Incised Text', *Pattern Recognition*, 36: 1031–43.
- Morik, K., S. Wrobel, J.-U. Kietz, and W. Emde (1993). *Knowledge Acquisition and Machine Learning: Theory, Methods and Applications* (London: Academic Press).
- Morwood, J., and M. Warman (1990). *Our Greek and Latin Roots* (Cambridge: Cambridge: University Press).
- Munby, L., S. Hobbs, and A. Crosby; (2002). *Reading Tudor and Stuart Handwriting*, 2nd edn., alphabet drawn by P. Judge (Chichester: British Association for Local History).
- Neal, R. M. (1993). 'Probabilistic Inference using Markov Chain Monte Carlo Methods', *Technical Report CRG-TR-93-1* (Toronto: Department of Computer Science, University of Toronto), <http://www.cs.toronto.edu/~radford/review.abstract.html> (accessed 20 Jan 2005).
- Norvig, P. (1991). *Paradigms of Artificial Intelligence Programming: Case Studies in Common LISP* (San Francisco: Morgan Kaufmann).
- Oakhill, J., and A. Garnham (1988). *Becoming a Skilled Reader* (Oxford: Basil Blackwell).

- Ogden, J. A. (1969). 'The Siting of Papyrus Fragments: An Experimental Application of Digital Computers', Ph.D. thesis, Department of Mathematics. Glasgow, University of Glasgow.
- Oostdijk, N. (1988). 'A Corpus Linguistic Approach to Linguistic Variation', *Literary and Linguistic Computing*, 3(1): 12–25.
- Oxford English Dictionary* (2005). Online: http://dictionary.oed.com/cgi/entry/50170837/50170837se5?single=1&query_type=word&queryword=papyrology&first=1&max_to_show=10&hilite=50170837se5 (accessed 3 Feb 2005).
- Pan, X., M. Brady, A. K. Bowman, C. Crowther, and R. S. O. Tomlin (2004). 'Enhancement and Feature Extraction for Images of Incised and Ink Texts', *Image and Vision Computing*, 22(6): 443–51.
- Parisse, C. (1996). 'Global Word Shape Processing in Off-Line Recognition of Handwriting', *Institute of Electrical and Electronics Engineers Transactions on Pattern Analysis and Machine Intelligence*, 18(4): 460–4.
- Parkes, M. B. (1991). *Scribes, Scripts and Readers* (London: Hambledon Press).
- (1992). *Pause and Effect: An Introduction to the History of Punctuation in the West* (Aldershot: Scolar Press).
- Pattie, T. S., and E. G. Turner (1974). *The Written Word on Papyrus* (London: British Museum Publications).
- Pestman, P. W. (1990). *The New Papyrological Primer, being the Fifth Edition of David and Van Groningen's Papyrological Primer* (Leiden: Brill).
- Plato (1986). *Phaedrus*, tr. C. J. Rowe (Warminster: Aris & Philips).
- Plautus (1965). 'Pseudolus', *The Pot of Gold and Other Plays*, tr. E. Watling (Harmondsworth: Penguin).
- Plunkett, K. (1995). 'Connectionist Approaches to Language Acquisition', in P. Fletcher and B. Macwhinney (eds.), *The Handbook of Child Language* (Oxford: Blackwell).
- Popping, R. (2000). *Computer Assisted Text Analysis* (London: Sage Publications).
- Prescott, A. (1997). 'The Electronic Beowulf and Digital Restoration', *Literary and Linguistic Computing*, 12(3): 185–95.
- Rabiner, L. R. (1989). 'A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition', *Proceedings of Institute of Electrical and Electronics Engineers*, 77(2): 257–86.
- Rayner, K. (1998). 'Eye Movements in Reading and Information Processing: 20 Years of Research', *Psychology Bulletin*, 124(3): 372–422.
- Reichgelt, H., and N. Shadbolt (1992). 'ProtoKEW: A Knowledge-Based System for Knowledge Acquisition', in D. Sleeman and N. Bersnen (eds.), *Artificial Intelligence*, vi. (Hove: Lawrence Erlbaum).

- Reiner, E. (1973). 'How we Read Cuneiform Texts', *Journal of Cuneiform Studies*, 25: 3–58.
- Renner, T. (1994). 'The Finds of Wooden Tablets from Campania and Dacia as Parallels to Archives of Documentary Papyri from Roman Egypt', in A. Bülow-Jacobsen (ed.), *Proceedings of the 20th International Congress of Papyrologists*, Copenhagen 23–29 Aug. 1992 (Copenhagen: Museum Tusculanum Press).
- Reynolds, L. D., and N. G. Wilson (1991). *Scribes and Scholars: A Guide to the Transmission of Greek and Latin Literature* (Oxford: Clarendon Press).
- Rissanen, J. (1978). 'Modeling by Shortest Data Description', *Automatica*, 14: 465–71.
- (1999). 'Hypothesis Selection and Testing by the MDL Principle', *Computer Journal*, 42(4): 260–9.
- Rivero, D., R. Vidal, J. Dorado, J. R. Rabuñal, and A. Pazos (2003). 'Restoration of Old Documents with Genetic Algorithms', *Proceedings of Applications of Evolutionary Computing: EvoWorkshops 2003*, EvoBIO, EvoCOP, EvoIASP, EvoMUSART, EvoROB, and EvoSTIM, Essex, UK, 14–16 Apr. 2003. (Lecture Notes in Computer Science 2611/2003; Heidelberg: Springer-Verlag, 432–43.
- Robertson, P. (1999). *A Corpus Based Approach to the Interpretation of Aerial Images*, Proceedings of 7th International Congress on Image Processing and its Applications (IEE IPA99) (Manchester), ii. 527–31.
- (2001). 'A Self Adaptive Architecture for Image Understanding', D.Phil thesis, Department of Engineering Science, University of Oxford.
- and R. Laddaga (2002). 'Principal Component Decomposition for Automatic Context Induction', in N. Ishii (ed.), *Artificial and Computational Intelligence: Proceedings of the IASTED International Conference* (Tokyo): 243–50.
- and R. Laddaga (2004). 'The GRAVA Self-Adaptive Architecture; History; Design; Application; and Challenges', *24th International Conference on Distributed Computing Systems Workshops (ICDCS 2004 Workshops)* (Institute of Electrical and Electronics Engineers Computer Society), 298–303.
- M. Terras, X. Pan, J. M. Brady, and A. K. Bowman (forthcoming). 'Image to Interpretation: An MDL Agent Architecture to Read Ancient Roman Texts', submitted.
- Rumelhart, D. E., and J. L. McClelland (1982). 'An Interactive Activation Model of Context Effects in Letter Perception: Part 2', *Psychological Review*, 89: 60–94.
- and — and the PDP Research Group (1986). *Parallel Distributed Processing, Explorations in the Microstructure of Cognition*, i. *Foundations*. (Cambridge, Mass.: MIT Press).

- Russell, S. J., and P. Norvig (1995). *Artificial Intelligence: A Modern Approach* (Upper Saddle River, N.J.: Prentice-Hall International).
- Saunders, L. (1999). 'The Uses of Greek', in W. Sieghart (ed.), *The Forward Book of Poetry 1999* (London: Forward Publishing), 53–7.
- Schenk, V. U. B. (2001). 'Visual Identification of Fine Surface Incisions', D.Phil. thesis, Department of Engineering Science, Oxford University.
- and M. Brady (2003). 'Visual Identification of Fine Surface Incisions in Incised Roman Stylus Tablets', *ICAPR 2003, International Conference in Advances in Pattern Recognition* (IEEE).
- Seales, W. B., J. Griffioen, K. Kiernan, C. J. Yuan, and L. Cantara (2000). 'The Digital Athenueum: New Technologies for Restoring and Preserving Old Documents', *Computers in Libraries*, 20(2): 26–50.
- Shadbolt, N., and M. A. Burton (1990). 'Knowledge Elicitation Techniques: Some Experimental Results', in K.L. McGraw, and C. R. Westphal (eds.), *Readings in Knowledge Acquisition, Current Practices and Trends* (London: Ellis Horwood), 17–23.
- Shaw, M. L. G., and B. R. Gaines (1983). 'A Computer Aid to Knowledge Engineering', *Proceedings of British Computer Society Conference on Expert Systems* (Cambridge: British Computer Society), 263–71.
- Sinclair, J. (1991). *Corpus, Concordance, Collocation* (Oxford: Oxford University Press).
- Smagorinsky, P. (1989). 'The Reliability and Validity of Protocol Analysis', *Written Communication*, 6(4): 463–77.
- Smith, F. (1971). *Understanding Reading: A Psycholinguistic Analysis of Reading and Learning to Read*. (New York: Holt, Rinehart & Winston).
- Solange-Pellat, E. (1927). *Les Lois d'écriture* (Paris: Vuibert).
- Stainton, R. J. (1996). *Philosophical Perspectives on Language* (Ontario: Broadview Press).
- Stark, J. A. (1992). *Digital Image Processing Techniques with Applications in Restoring Ancient Manuscripts* (Cambridge: Department of Engineering Science, University of Cambridge, EIST Project Report).
- Stemler, S. (2001). 'An Overview of Content Analysis', *Practical Assessment, Research and Evaluation*, 7(17), <http://pareonline.net/getvn.asp?v=7&n=17> (accessed 27 Jan 2005).
- Stewart, V. (2004). 'Business Applications of Repertory Grid', http://www.enquirewithin.co.nz/BUS_APP/business.htm (accessed 13 Jan 2005).
- Stoppard, T. (1997). *The Invention of Love* (London: Faber & Faber).
- Stubbs, M. (1996). *Text and Corpus Analysis: Computer Assisted Studies of Language and Culture* (Oxford: Blackwell).

- Swinney, D. A., and D. T. Hakes (1976). 'Effects of Prior Context upon Lexical Access during Sentence Comprehension', *Journal of Verbal Learning and Verbal Behaviour*, 15(6): 681–9.
- Tacitus, C. (1964). *Tacitus on Britain and Germany: A Translation of the Agricola and the Germania*, tr. H. Mattingly (Harmondsworth: Penguin Books).
- Terras, M. (2002). 'Image to Interpretation: Towards an Intelligent System to Aid Historians in the Reading of the Vindolanda Texts', D.Phil. thesis, Department of Engineering Science, University of Oxford.
- and P. Robertson (2004). 'Downs and Acrosses: Textual Markup on a Stroke Based Level', *Literary and Linguistic Computing*, 19(3): 397–414.
- and —— (2005). 'Image and Interpretation: Using Artificial Intelligence to Read Ancient Roman Texts', *HumanIT*, 7: 3.
- Thomas, J. D. (1976). 'New Light on Early Latin Writing', *Scriptorium*, 30: 38–43.
- (1992). 'The Latin Writing Tablets from Vindolanda in North Britain', in E. Lalou, (ed.), *Bibliologia: Les Tablettes à écrire de l'antiquité à l'époque moderne*. (Turnhout: Brepols), 12: 204–7.
- Thompson, E. M. (1912). *Introduction to Greek and Latin Palaeography* (Oxford: Clarendon Press).
- Thoyts, E. E. (1893). *How to Decipher and Study Old Documents: Being a Guide to the Reading of Ancient Manuscripts* (London: Elliot Stock).
- Tjäder, J. O. (1977). 'Latin Palaeography, 1975–7', *Erano*, 75: 131–60.
- (1979). *Considerazione e proposte sulla scrittura Latina nell' età Romana* (Rome: Edizioni di Storia e Letteratura).
- (1986). 'Review of Tabulae Vindolandenses I', *Scriptorium*, 40: 297–301.
- Tomlin, R. S. O. (1985). 'The Inscribed Objects', in P. T. Bidwell (ed.), *The Roman Fort of Vindolanda at Chesterholm, Northumberland* (London: Historic and Monuments Commission for England) ch.17.
- (1988). 'The Finds from the Sacred Springs', in B. Cunliffe (ed.), *The Temple of Sulis Minerva at Bath, VII*. (Oxford: Oxford University Committee for Archaeology, Monograph 16); 4–277.
- (1994). 'Vinisius to Nigra: Evidence from Oxford of Christianity in Roman Britain', *Zeitschrift für Papyrologie und Epigraphik*, 100: 93–108.
- (1996). 'Review Article: The Vindolanda Tablets', *Britannia*, 27: 459–63.
- Tunmer, W. E., and W. A. Hoover (1992). 'Cognitive and Linguistic Factors in Learning to Read', in P. B. Gough, L. C. Ehri, and T. Rebecca (eds.), *Reading Acquisition* (London: Lawrence Erlbaum Associates), 175–214.
- Turner, E. G. (1968). *Greek Papyri: An Introduction* (Oxford: Clarendon Press).
- *The Papyrologist at Work* (Durham, N.C.: Duke University).

- Turner, E. G. (1980). *Greek Papyri: An Introduction*, rev. edn. (Oxford: Clarendon Press).
- Ullman, B. L. (1969). *Ancient Writing and its Influence* (Cambridge, Mass.: MIT Press).
- Underwood, G., and R. Radach (1998). 'Eye Guidance and Visual Information Processing: Reading, Visual Search, Picture Perception and Driving', in E. Underwood (ed.), *Eye Guidance in Reading and Scene Perception*. (Oxford: Elsevier), 1–28.
- van Groningen, B. A. (1932). 'Project d'unification des systèmes des signes critiques', *Chronique d'Égypte*, 7: 262–9.
- van Hoesen, H. B. (1915). *Roman Cursive Writing* (Princeton: Princeton University Press).
- van Minnen, P. (1995). 'The Century of Papyrology (1892–1992)', <http://odyssey.lib.duke.edu/papyrus/texts/history.html> (accessed 27 Jan 2005).
- Venuti, L. (ed.) (2000). *The Translation Studies Reader* (London: Routledge).
- Viterbi, A. J. (1967). 'Error Bounds for Convolution Codes and an Asymptotically Optimal Decoding Algorithm', *Institute of Electrical and Electronics Engineers Transactions on Information Theory*, 13: 260–9.
- Wacholder, B.-Z., and M. G. Abegg (1991). *A Preliminary Edition of the Unpublished Dead Sea Scrolls, etc.* fascicle 1. (Washington, D.C.: Biblical Archaeological Society).
- Wallace, C. S., and D. M. Boulton (1968). 'An Information Measure for Classification', *Computer Journal*, 11: 185–95.
- Ware, G. A., D. M. Chabries, R. W. Christiansen, and C. E. Martin (2000). 'Multispectral Document Enhancement: Ancient Carbonized Scrolls', *Proceedings Institute of Electrical and Electronics Engineers 2000 International Geoscience and Remote Sensing Symposium*, (IEEE), vi. 2486–8.
- Waterman, D. A. (1986). *A Guide to Expert Systems* (Reading, Mass.: Addison-Wesley).
- Weinreich, M. (1945). 'YIVO and the Problems of our Time', *Yivo-Bleter*, 25(1): 10–24.
- White, S. (2000). 'Enhancing Knowledge Acquisition with Constraint Technology', D.Phil. thesis, Department of Computer Science, University of Aberdeen.
- Winter, J. G. (1936). 'Papyrology: Its Contributions and Problems', *Michigan Alumnus Quarterly Review* (Summer): 234–48.
- Woolley, S. I., N. J. Flowers, T. N. Arvanitis, A. Livingstone, T. R. Davis, and J. Ellison (2001). '3D Capture, Representation and Manipulation of Cuneiform Tablets', *IST/SPIE Electronic Imaging 2001*, Proc SPIE (Three Dimensional Image Capture and Applications IV), 4298: 103–10.

- Wordsworth, W. (1819). 'Upon the Same Occasion', in W. Knight (ed.), *The Poetical Works of William Wordsworth*, vi (Edinburgh: William Paterson, 1884), 182–5.
- Yarbus, A. L. (1967). *Eye Movements and Vision* (New York: Plenum Press).
- Youtie, H. C. (1963). 'The Papyrologist: Artificer of Fact', *Greek, Roman and Byzantine Studies*, 4: 19–32.
- (1966). 'Text and Context in Transcribing Papyri', *Greek, Roman and Byzantine Studies*, 7: 251–8.
- Zhu, S. C., and A. Yuille (1996). 'Region Competition: Unifying Snakes, Region Growing, and Bayes/MDL for Multi-Band Image Segmentation', *Institute of Electrical and Electronics Engineers Transactions on Pattern Analysis and Machine Intelligence*, 18(9): 884–900.
- Zhang, Y., and K. K. Tan (2004). *Convergence Analysis of Recurrent Neural Networks* (Network Theory and Applications, 13; Boston: Kluwer Academic Publishers).
- Zipf, G. K. (1935). *The Psycho-Biology of Language* (Cambridge, Mass.: MIT Press, repr. 1965)
- Zuckerman, B. (1996). 'Bringing the Dead Sea Scrolls Back to Life: A New Evaluation of Photographic and Electronic Imaging of the Dead Sea Scrolls', *Dead Sea Discoveries*, 3(2): 178–207.

This page intentionally left blank

Index

- Aalto, P. 31, 55–6
agents 58, 123, 124, 124 n. 1, 125, 126,
167, 173, 184–5 *see* GRAVA
Agricola 2
artificial intelligence 15, 16, 19, 72, 91,
106, 113, 119, 124 n. 1, 124–5,
125 n. 2, 167, 170, 171, 175, 180
automatic feature extraction 21, 126,
130–2, 148, 151, 163–4, 166, 169, 187

Bayes' Theorem 194
Bayesian inference 178 n. 10
bilingualism 155
blackboard system 177 n. 7
Bowman, A. K. 32–4, 56, 72, 87, 91

Capitalis 86, 155
character models *see* image annotation,
character models
Classical Latin 112, 161
cognitive visual system 166, 168
content analysis 47–50, 61
definition and history 48
encoding scheme 49, 57, 75–7
validity 50
convergence analysis 142 n. 1, 166, 173,
180
corpus of annotated images, *see* image
annotation
Cuneiform 29, 31, 155

description length 175, 176, 194, 197 *see*
Minimum Description Length
digitization 13, 168

encoding scheme *see* image annotation
EpiDoc 13 n. 27,
eXtensible Markup Language *see* XML
eye movements in reading 83, 154, 157–8

facial recognition 169
forensic palaeography *see* palaeography
forward chaining processing 177 n. 8

GRAVA (Grounded Reflective Adaptive
Vision Architecture) 2, 20, 21, 124,
125, 126–9, 133–7, 138, 153
adapted version 133, 138–9
agent behaviour 166–7, 173–6,
181–2, 184–5
annotator program 105–7, 131, 187
character agent 126, 133–4, 135–7,
150, 151, 171, 174, 188, 193–6,
computational efficiency 166–7, 184,
186
future development 150, 156–7,
158–65, 168–9
results 138–152
screen output 130
system architecture 23, 126, 131,
133–5, 169, 171–2, 173, 180–5, 187
training 126
use of Minimum Description
Length 126–7, 134, 135–7, 182, 186
use of Monte Carlo methods 171,
184–5, 196 *see* Monte Carlo
methods
word agent 126, 133–4, 150, 174,
183–4, 188, 197–9
grammar 31, 44, 49, 53, 63, 72, 76, 112,
169
modelling 160–1, 167, 169, 178
of Latin 72, 112, 160–1, 178

Hadrian's Wall 2–3
handwriting recognition techniques
18–19, 169 *see* GRAVA
heterarchical programming
approaches 175 n. 6
Hidden Markov Models 167, 178–9,
Hunt, A. S. 33, 34, 56

image annotation 91, 105
annotated corpus 24, 84–5, 92,
105–6, 118, 123, 138, 156, 186,
207–10, 211
automatic annotation 130–2, 133

- image annotation (*cont.*)
- character models 117, 132–3, 139, 145, 150, 156, 158, 162, 167, 171–2, 186–7, 188–93
 - encoding scheme 21, 23, 24, 85, 90–1, 92, 97–100, 118, 132, 167–8
 - example 106–9, 111, 191, 205
 - methodology 105, 107–8, 110, 118–19, 200
 - textual annotation 109, 162
 - see* GRAVA annotator program
- image processing 9–12, 13, 15, 158, 163, 164, 166, 169, 186
- automatic feature extraction, *see* automatic feature extraction
 - applied to ancient texts 13–14
 - Gaussian blur operator 188, 189–91
 - integration with AI 15, *see* minimum description length
 - medical image analysis 169
 - minimum description length, *see* minimum description length
 - phase congruency *see* phase congruency
 - wavelet filtering technique 10
- infrared photography 1, 6–7
- ink texts 1, 3, 5, 6–7
- date of creation 5
 - discovery 5–6,
 - examples 7, 45, 46, 101–4
 - expert behaviour in reading 65–8, 74
 - hands 88, 101–4
 - letter forms 84, 87, 115–17
 - physical characteristics 6
 - preservation 5
 - subject matter 7, 101–4
 - test set for GRAVA system 139–141, 143–145, 163
- Interactive Activation and Competition Model of Word Recognition, *see* McClelland and Rumelhart
- interpretation problem 19, 168–9, 170, 173, 176, 178
- James, M. R. 37
- Java 163
- knowledge acquisition 38, 42
- knowledge engineer *see* knowledge elicitation
- knowledge elicitation 14, 16, 17, 23, 27, 38–41, 84, 91–3, 118, 154–8, 167–8
- automated 39–40, 51
 - history 39,
 - protocols 39, 41, 44 *see* Think Aloud Protocols *see* Repertory Grid
- Leiden system 33, 42, 70–2
- symbols 70 n. 1
- letter forms 23, 32, 36–7, 65, 73, 84, 92, 94–7, 98–100, 156, 192–3, 211
- see* image annotation, character models
 - see* Old Roman Cursive
- lexicostatistics 21, 84, 85, 111–14, 118, 123, 174
- corpus size and validity 113–14
 - expanding current corpus 150, 154, 158–162
 - letter frequency 113, 114, 138, 144, 159, 167, 168
 - word list 111–13, 138, 159–62, 167, 168
- Lisp 125 n.1, 163, 173, 182, 185
- markup 13, 98, 106, 119
- see* image annotation
 - see* XML
- Mathyer, J. 89–90
- McClelland and Rumelhart viii, 17, 20, 79–81
- Interactive Activation and Competition Model of Word Recognition 79–80, 125, 172
- minimum description length 2, 17, 19–21, 22, 23, 58, 124, 125, 127, 137, 140–3, 149, 151, 153–4, 166, 171, 172, 175, 176–7, 186, 194–5, 197
- relation to probability (P) 175
 - use in image processing 20, 126–9, 139, 149, 151, 166, 168–9
- Monte Carlo methods 20–1, 23–4, 142, 166–7, 171, 177–80, 179, 186, 199
- applied to agent selection 171, 179, 184–5
 - definition 20–21 n. 36
- Markov Chain Monte Carlo Methods 178 n. 10
- reasons for use 179

- Motif 105 n. 12
 multi-spectral imaging 14
- natural language processing 91, 167, 178–9
- New Roman Cursive 87, 155
- Occam's razor 19–20
- oculomotor action *see* eye movements
 in reading
- Old Roman Cursive 84, 118, 168
 definition 86
 extant sources 86 n. 3, 87 n. 5, 91
 difference in forms on ink and stylus
 texts 85, 88–91, 115–17, 145, 168, 190, 192–3
 letter forms 86–7, 88, 89, 91–2, 93, 94–7, 115–17, 190, 192–3,
 stroke order 96
 use in Roman Britain 87
- palaeography
 definition 85–6
 forensic palaeography 89–91
 of Vindolanda *see* Old Roman Cursive
- papyrology
 army intelligence 30–1
 computational model 58, 123, 125, 153
 critical apparatus 42–3, 50, 59–61, 68, 69
 critical discussions 29–34
 definition 12, 28
 established computational methods
 and projects 12–14, 168
 expert behaviour and methods 9, 11, 16, 27–83, 153–5, 164–5, 167–8
 expert system 22, 163–5
 history 28–9
 model of process 17, 81–2, 83, 123, 125, 167–8
 recursive nature of process 55–58, 75, 156–7
 process of reading 15, 17, 28–34, 37, 38, 41, 43, 72–4, 164–5,
 subject matter for art 33
- Parallel Distributed Processing 172–3, 172 n. 1, 175 n. 6
- Perseus Digital Library 13, 113–14
- phase congruency 10–12, 129, 130, 158, 169
- PhotoShop 12, 13, 165
- Plautus, *Pseudolus* (1.1.23–30) 9
- Plato, *Phaedrus* (274–5) 123
- psycholinguistics 34–6, 167
 lexical ambiguity 36
 structural ambiguity 36
- rapid prototyping 52–3 n.21
- reading 36–7, 77, 77–9, 157
 computational modelling 78–9, 80, 173 *see* McClelland and Rumelhart
- Reiner, E. 1, 31, 55
- Repertory Grid 39, 40 n. 12, 93–4
 WebGrid 93–4, 156
- Robertson, P. 105–6, 124–9, 172–6, 185–6
- Rumelhart and McClelland, *see* McClelland and Rumelhart
- self-adaptive architecture 127
- shadow stereo, *see* phase congruency
- sign language interpretation 169
- Stanegate 2–3
- stylus tablets 1, 3, 5, 7–9, 32
 date of creation 5
 discovery 5–6
 examples 8, 42, 47, 48
 expert behaviour in reading 65–8, 74
 letter forms 84, 87, 88–91, 105, 115–17
 physical characteristics 7–9
 preservation 5
 subject matter 8, 42, 104
 test set for GRAVA system 145–8, 163
- subsumption 177–8 n. 9
- Stochastic Context Free Grammars 178 n. 12
- Tacitus 5
- Textual Encoding Initiative (TEI) 13 n. 27, 109
- textual analysis 50–1, 67, 71, 150, 161–2
 see lexicostatistics
- Thomas, J. D. 87, 91
- Thoyts, E. 31–2, 56

- Think Aloud Protocols 39, 43, 44–5,
50, 59, 65, 72, 83, 92
methodology with Vindolanda
experts 45–7, 93, 155–6
- Tomlin, R. 32–3, 56, 72
- Turner, E. G. 33, 56
- translation studies 35
- Vindolanda
archaeological excavations 5–6
discovery of tablets 5–6
geography 1, 4
fort 1–4
ink texts, *see* ink texts
occupation 2–3, 5
stylus tablets, *see* stylus tablets
- Tabulae Vindolandenses* 42–3, 50,
59–61, 68, 69, 71, 91–2, 112,
vicus 3
- Viterbi Algorithm 178 n. 13, 179
- Vulgar Latin 111–12, 112 n. 18, 114
- WebGrid *see* Repertory Grid
- XML (eXtensible Markup
Language) 13 n. 27, 85, 106,
107–9, 118–19, 131, 162, 163,
204
- Yolambda 125 n. 1, 163, 173
- Youtie, H. C. 29–30, 38, 55–6